



PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books
Search PubMed for [] Go Clear
☒ Limits Preview/Index History Clipboard Details

About Entrez

Display Abstract Show: 20 Sort Send to Text

Text Version

☐ 1: Transgenic Res 2002 Aug;11(4):425-35

[Related Articles, Links](#)

Entrez PubMed

Overview
Help | FAQ
Tutorial
New/Noteworthy
E-Utilities

PubMed Services

Journals Database
MeSH Browser
Single Citation Matcher
Batch Citation Matcher
Clinical Queries
LinkOut
Cubby

Related Resources

Order Documents
NLM Gateway
TOXNET
Consumer Health
Clinical Alerts
ClinicalTrials.gov
PubMed Central

Privacy Policy

The use of the uromodulin promoter to target production of recombinant proteins into urine of transgenic animals.

Zbikowska HM, Soukhareva N, Behnam R, Chang R, Drews R, Lubon H, Hammond D, Soukharev S.

Plasma Derivative Department, Holland Laboratory, American Red Cross, MD 20855, Rockville, USA.

A uromodulin promoter has been isolated, sequenced, and used to generate two sets of transgenic mice for expression of the lacZ marker gene and for production of the human recombinant erythropoietin (rhEPO) in urine. We demonstrated that the 5.6-kb fragment of the uromodulin gene containing the 3.7-kb promoter area and, both the first exon and part of the second exon, were sufficient to provide kidney-specific expression of the lacZ gene. Histological analysis of the lacZ expression pattern revealed beta-galactosidase activity specifically in the thick limb of Henle's loop. However, due to random integration of the transgene, ectopic expression was detected in some transgenic lines. Analysis of the EPO-transgenic mice showed that rhEPO was secreted into the urine of founder mice (up to 6 ng/ml). We were able to breed and analyze only two sublines with a very low expression level of rhEPO (up to 260 pg/ml). All of our transgenic mice expressing rhEPO in urine developed disease symptoms similar to polycythemia in humans. These included a considerable increase in red blood cell counts, hemoglobin concentration, and hematocrit concomitant with severe thrombocytopenia, all of which were detected in the rhEPO-expressing mice. Although our model did not prove to be beneficial for commercial production of rhEPO, we concluded that the uromodulin promoter could be useful for expression of other important therapeutic proteins into the urine of transgenic animals.

PMID: 12212844 [PubMed - in process]

Display Abstract Show: 20 Sort Send to Text

Write to the Help Desk
NCBI | NLM | NIH
Department of Health & Human Services
Freedom of Information Act | Disclaimer

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 17, 2003, 08:54:07 ; Search time 54.8618 Seconds
(without alignments)
9111.676 Million cell updates/sec

Title: US-09-605-042A-37

Perfect score: 1630
Sequence: 1 acataggagcagcggtgtc.....ccattctgtgaagatg 1630

Scoring table: IDENTITY_MNC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/ptodata/2/1na/5A.COMB.seq:*
- 2: /cgn2_6/ptodata/2/1na/5B.COMB.seq:*
- 3: /cgn2_6/ptodata/2/1na/6A.COMB.seq:*
- 4: /cgn2_6/ptodata/2/1na/6B.COMB.seq:*
- 5: /cgn2_6/ptodata/2/1na/PCFUS.COMB.seq:*
- 6: /cgn2_6/ptodata/2/1na/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	47.2	2.9	7218	1 US-08-232-463-14	Sequence 14, Appl
C 2	41.4	2.5	1001	4 US-09-641-638-161	Sequence 161, App
C 3	41	2.5	5064	4 US-08-936-107A-8	Sequence 8, Appl
4	40.2	2.5	565	4 US-09-323-195A-5	Sequence 5, Appl
5	40	2.5	340	4 US-08-323-195A-3	Sequence 3, Appl
6	39.4	2.4	240	4 US-08-628-417-6	Sequence 6, Appl
7	39	2.4	341	4 US-09-323-195A-1	Sequence 1, Appl
C 8	39	2.4	3172	4 US-08-978-289-11	Sequence 11, Appl
9	38.8	2.4	7218	1 US-08-232-463-14	Sequence 14, Appl
10	38.4	2.4	3792	2 US-08-992-334-1	Sequence 1, Appl
11	38.4	2.4	5234	2 US-08-302-752-1	Sequence 1, Appl
12	38.4	2.4	5234	2 US-08-992-334-2	Sequence 2, Appl
13	38.4	2.4	6722	2 US-08-992-334-3	Sequence 3, Appl
14	38.4	2.4	6722	2 US-08-302-752-3	Sequence 3, Appl
15	38.4	2.4	6722	2 US-08-302-752-3	Sequence 3, Appl
16	38.2	2.3	2791	4 US-09-570-367C-1	Sequence 1, Appl
17	37.8	2.3	1798	4 US-09-797-906-1	Sequence 1, Appl
18	37.8	2.3	19124	2 US-08-487-826B-13	Sequence 13, Appl
C 19	37.6	2.3	3138	1 US-07-867-106-4	Sequence 4, Appl
C 20	37.4	2.3	6078	1 US-09-173-914-1	Sequence 1, Appl
C 21	37.2	2.3	1924	4 US-09-424-283-5	Sequence 5, Appl
C 22	37	2.3	342	4 US-09-323-195A-6	Sequence 6, Appl
C 23	37	2.3	658	4 US-08-998-416-595	Sequence 595, App
C 24	37	2.3	921	4 US-09-377-648-4	Sequence 4, Appl
25	37	2.3	1298	4 US-08-971-089-1	Sequence 1, Appl
26	37	2.3	4526	4 US-09-424-283-7	Sequence 7, Appl
27	36.8	2.3	868	4 US-09-071-035-27	Sequence 27, Appl

28	36.8	2.3	960	4 US-09-071-035-25	Sequence 25, Appl
29	36.8	2.3	2223	1 US-08-257-073-4	Sequence 4, Appl
30	36.8	2.3	2447	2 US-09-014-969-14	Sequence 14, Appl
C 31	36.6	2.2	255	4 US-09-257-583-5	Sequence 5, Appl
C 32	36.6	2.2	731	1 US-08-451-405A-2	Sequence 2, Appl
C 33	36.2	2.2	1733	3 US-09-073-569-1	Sequence 1, Appl
34	36	2.2	48	4 US-08-913-014A-18	Sequence 18, Appl
C 35	36	2.2	336	4 US-09-276-599-13	Sequence 13, Appl
C 36	36	2.2	763	4 US-09-276-599-14	Sequence 14, Appl
37	36	2.2	1447	4 US-09-443-041A-27	Sequence 27, Appl
38	36	2.2	1700	2 US-08-897-340-4	Sequence 4, Appl
39	36	2.2	1700	3 US-09-252-329-4	Sequence 4, Appl
C 40	36	2.2	2791	4 US-09-570-367C-1	Sequence 1, Appl
41	36	2.2	3718	4 US-09-424-283-6	Sequence 6, Appl
42	36	2.2	37948	4 US-09-251-645-11	Sequence 11, Appl
43	35.8	2.2	621	4 US-09-328-111-177	Sequence 177, App
44	35.8	2.2	3275	4 US-09-370-838-151	Sequence 151, App
45	35.8	2.2	6755	3 US-08-931-999-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESSES:
ADDRESSSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935, 313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PTZ9pt-Fls
US-08-232-463-14
Query Match 2.9%; Score 47.2; DB 1; Length 7218;

```
Best Local Similarity   3.5%; Pred. No. 0.00081; Indels      0; Gaps      0
Matches    13; Conservative 210; Mismatches 153;
```

RESULT 2
US-09-641-638-161/c
US-09641638

; Sequence ID: 6432848
; Patent No. 6432848
; GENERAL INFORMATION:
APPLICANT: Blumenfeld, Marta
APPLICANT: Bouguenerec, Lydie
APPLICANT: Chumakov, Ilya
APPLICANT: Cohen, Amick
TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
TITLE OF INVENTION: GENSET.05ICP1
FILE REFERENCE: GENSET.05ICP1
CURRENT FILING DATE: US/09/641,638
CURRENT FILING DATE: 2000-08-16
PRIOR APPLICATION NUMBER: US 09/502,330
PRIOR FILING DATE: 2000-02-11
PRIOR APPLICATION NUMBER: US 60/133,200
PRIOR FILING DATE: 1999-05-07
PRIOR APPLICATION NUMBER: US 09/275,267
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: US 60/119,917
PRIOR FILING DATE: 1999-02-12
NUMBER OF SEQ ID NOS: 1304
SOFTWARE: Patent.pm
SEQ ID NO: 161
LENGTH: 1001
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: allele
LOCATION INFORMATION: 501
OTHER INFORMATION: 12-747-302 : polymorphic base C or T
NAME/KEY: misc_binding
LOCATION: 481..500
OTHER INFORMATION: 12-747-302.misl, potential
NAME/KEY: misc_binding
LOCATION: 502..521
OTHER INFORMATION: 12-747-302.mis2, potential complement
NAME/KEY: primer_bind
LOCATION: 200..219
OTHER INFORMATION: upstream amplification primer

Query Match	2.58%	Pred. No. 0.014%	Indels	0	Gaps
Best Local Similarity	50.8%				
Matches	99	Mismatches	96		
					143

[illegible]

RESULT 3
TC-08-936-107A-8
TC-08-936-107A-8

Sequence 8 Application US/00/000000
Patent No. 6403306
GENERAL INFORMATION:
Applicant: Stephens, David S.
Applicant: Swartley, John S.
TITLE OF INVENTION: Serogroup-Specific Nucleotide Sequences
TITLE OF INVENTION: in the Molecular Typing of Bacterial Isolates and the
TITLE OF INVENTION: Preparation of Vaccines Thereof
TITLE OF INVENTION: 51
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
STREET: 5370 Manhattan Circle, Suite 201
City: Boulder
STATE: Colorado
COUNTRY: US
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/936,107A
FILING DATE: 23-SEP-1997
CLASSIFICATION: 536
PRIORITY APPLICATION DATA:
PRIORITY APPLICATION NUMBER: US 08/827,622
FILING DATE: 09-APR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Caruthers, Jennie M.
REGISTRATION NUMBER: 34,464
REFERENCE/DOCKET NUMBER: 77-97
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 499-8080
TELEFAX: (303) 499-8089
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 5064 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: not relevant
MOLECULE TYPE: DNA (genomic)

Wed Feb 19 14:09:23 2003

us-09-605-042a-37.rml

HYPOTHETICAL: NO
 FEATURE: CDS
 NAME/KEY: 479..1597
 LOCATION: 1599..3236
 FEATURE: CDS
 NAME/KEY: 3309..4052
 LOCATION: 4054..4917
 NAME/KEY: CDS
 LOCATION: 4054..4917
 US-08-936-107A-8

2.5%: Score 41; DB 4; Length 5064;
 Query Match
 Best Local Similarity 47.58; 0; Mismatches 135; Indels 0; Gaps 0;
 Matches 122; Conservative

76 CACCAACGAAATGACATTCACGACCTAGGATGAGAAAAGAAATATTAGGACTTTA 135
 1537 CCCTTAGGGATGTATAGCTCCTCAAAATCATGTGTAGAACTTTATTAAGATAGATA 1596
 136 TTTCTTCTGATTTATAGCAAGAAAGGAGAAACCTTTATTAAGATAGATA 1596
 1597 AAATTTTATCTTATATACGAAATGCGCTTAACTTTAAAGATAGATA 1596
 196 ACGGCAAGATCAACACAGTTCAGAAAAGATTAATAGTATAGATTTCTCACTTTCA 315
 1657 TTGAGATAGTAAATTTAACTTTTATAGATTTTCTGCTAAATTTGCAAGAAAT 1716
 256 TTCTTTTATGCAAGACGCAAAAGTAAATATAGTATAGATTTCTCACTTTCA 315
 1717 TTAATAATTCATCATTCATATCCATAAACTAATATATAGTAAAGCTCAATCAATTTTCT 1776
 316 CAAGATGAAAGTTAT 332
 1777 CAACCTAAAGAAAT 1793

RESULT 4
 US-09-323-195A-5
 Sequence 5, Application US/09323195A

GENERAL INFORMATION:
 APPLICANT: Pullman, John
 APPLICANT: Caltrera, Juan
 APPLICANT: Petreia, Juan
 TITLE OF INVENTION: VITILIN-LIKE SEED STORAGE PROTEIN GENE PROMOTER AND
 TITLE OF INVENTION: METHODS OF USING THE SAME
 FILE REFERENCE: IPST0009
 CURRENT FILING DATE: 1999-06-01
 CURRENT FILING DATE: 1999-06-01
 NUMBER OF SEQ ID NOS: 19
 SOFTWARE: Patentln Ver. 2.0

SEQUENCE: 565
 LENGTH: 565
 TYPE: DNA
 ORGANISM: Pinus taeda
 FEATURE: n at 489 is a, c, g, or t
 OTHER INFORMATION: n at 503 is a, c, g, or t
 OTHER INFORMATION: n at 504 is a, c, g, or t
 OTHER INFORMATION: n at 522 is a, c, g, or t
 OTHER INFORMATION: n at 533 is a, c, g, or t
 OTHER INFORMATION: n at 543 is a, c, g, or t
 OTHER INFORMATION: n at 549 is a, c, g, or t
 OTHER INFORMATION: n at 564 is a, c, g, or t
 OTHER INFORMATION: n at 564 is a, c, g, or t
 US-09-323-195A-5

Query Match
 Best Local Similarity 84.98; 0; Mismatches 8; Indels 0; Gaps 0;
 Matches 45; Conservative

US-09-323-195A-3
 Sequence 3, Application US/09323195A

GENERAL INFORMATION:
 APPLICANT: Pullman, John
 APPLICANT: Caltrera, Juan
 APPLICANT: Petreia, Juan
 TITLE OF INVENTION: VITILIN-LIKE SEED STORAGE PROTEIN GENE PROMOTER AND
 TITLE OF INVENTION: METHODS OF USING THE SAME
 FILE REFERENCE: IPST0009
 CURRENT FILING DATE: 1999-06-01
 CURRENT FILING DATE: 1999-06-01
 NUMBER OF SEQ ID NOS: 19
 SOFTWARE: Patentln Ver. 2.0

2.5%: Score 40; DB 4; Length 340;
 Query Match
 Best Local Similarity 100.0%; 0; Mismatches 0; Indels 0; Gaps 0;
 Matches 40; Conservative

1 ACTATAGGCAAGCGGTGTCAGCGCCCGGCTGTTAAAGACACCCAGACTTA 53
 1 ACTATAGGCAAGCGGTGTCAGCGCCCGGCTGTTAAAGATTTACTTACTTA 53

RESULT 6
 US-08-628-417-6
 Sequence 6, Application US/08628417

GENERAL INFORMATION:
 APPLICANT: GILLESPIE, DAVID
 APPLICANT: COMPETITOR PRIMER ASYMMETRIC
 TITLE OF INVENTION: POLYMERASE CHAIN REACTION
 TITLE OF INVENTION: POLYMERASE CHAIN REACTION
 NUMBER OF SEQUENCES: 7
 CORRESPONDENCE ADDRESS: CHEMICAL AND BIOLOGICAL
 ADDRESS: U.S. ARMY
 ADDRESS: DEFENSE COMMAND
 STREET: OFFICE OF THE CHIEF COUNSEL (AMSCB-GC)
 CITY: ABERDEEN PROVING GROUND
 STATE: MARYLAND
 COUNTRY: USA
 ZIP: 21010-5423

COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentln Release #1.0, Version #1.25
 CURRENT APPLICATION NUMBER: US/08/628,417
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: BIFONTI, ULYSSES J
 REGISTRATION NUMBER: 39,908
 TELEPHONE: 410-671-1158
 TELEFAX: 410-671-2534
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 240 bases
 TYPE: nucleic acid


```

ANTI-917-6
US-08-628-817-6
Query Match          2.4%; Score 39.4; DB 1; Length 240;
Best Local Similarity 50.3%; Fred. No. 0.025; Indels 0; Gaps
Matches 97; Conservative 0; Mismatches 96;
OY 106 GATTGAGAAAAAGATTTATTTGACAACTTTATTCTTCTGAGCTATATACAGAAAGAAAGCG 64
    111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 5 GATTAAGACAAATAAADACTTTAGAAAATATTTACTATATATATATATATATATATATATATATAT 124
    111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
OY 166 GATAAAAAAAAACATCTTTATGCGGATTAACCGCAAGCTACCAACGTTCCAGAAA 225
    111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 65 AAAAAAATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 124
    111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
OY 226 GATTAATATCTAGCAATGAAAAGATTAATCTCTCTTTTCATCAAGAACGCCAAAAGTA 285
    111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 125 AAAAAAATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 184
    111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
OY 286 AATTAATGATAGA 298
    111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 185 AAAAAAAAAAAAAA 197
    111 111 111 111 111 111 111 111 111 111 111 111 111 111 111

```

RESULT 7
MS-09-323-195A-1
US/09323195A

```

1  GENERAL INFORMATION: Gerald
2  APPLICANT: Pullman, John
3  APPLICANT: Catriney, Rathan
4  APPLICANT: Pereira, Rathan
5  TITLE OF INVENTION: VICILIN-LIKE SEED STORAGE PROTEIN GENE PROMOTER AND
6  TITLE OF INVENTION: METHODS OF USING THE SAME
7  FILE REFERENCE: IPST0009
8  CURRENT APPLICATION NUMBER: US/09/323,195A
9  CURRENT FILING DATE: 1999-06-01
10 NUMBER OF SEQ ID NOS: 19
11 SOFTWARE: PatentIn Ver. 2.0
12 SEQ ID NO 1
13 LENGTH: 341
14 TYPE: DNA
15 ORGANISM: Pinus taeda
16 OS-09-323-195A-1

```

Organism	Query Match	Best Local Match	Score	DB	Length	Indels	Gaps
US-09-323-195A-1	2.4%	100.0%	39	0.04	0	0	0
Matches	39	Conservative	39				
QY	1	ACTATAGGCGACGCGTGTCTGACGCGCCGGCTGTATAA	39				
Db	1	ACTATAGGCGACGCGTGTCTGACGCGCCGGCTGTATAA	39				

RESULT 8
 US-08-978-289-11/C
 Sequence 11, Application US/08978289
 Patent No. 6361939
 GENERAL INFORMATION: Elizabeth E.M.
 APPLICANT: Bates, Blandine M.
 APPLICANT: Caux, Christophe
 APPLICANT: Lebecque, Serge J.E.
 APPLICANT: Banchereau jacques
 TITLE OF INVENTION: ISOLATED MAMMALIAN DENDRITIC CELL GENES
 TITLE OF INVENTION: RELATED REAGENTS
 NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:
 ADDRESSEE DNAX Research Institute

STREET: 901 California Avenue
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1104
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patentIn Release #1.0, Version #1.30
CURRENT APPLICATION NUMBER: US/08/978, 289
FILING DATE: 25-NOV-1997
536

CDR APPLICATION DATA: US 60/032,767
 PRIOR APPLICATION NUMBER:
 FILING DATE: 11 DEC-1996
 PRIOR APPLICATION DATA: US 60/031,806
 APPLICATION NUMBER: 27-NOV-1996
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Ching, Edwin P. 34, 090
 REGISTRATION/DOCKET NUMBER: DX0669KX
 REFERENCE/COMMUNICATION INFORMATION:
 TELEPHONE: (650)857-9196
 TELEFAX: (650)456-1204
 TELETYPE FOR SEQ ID NO: 11:
 INFORMATION CHARACTERISTICS:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3172 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 ORIGIN: DNA
 SEQUENCE TYPE: CDS

```

NAME/KEY: CDS..1290
LOCATION: 43..1290
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 112..1290
FEATURE:
NAME/KEY: misc_feature
LOCATION: 481
OTHER INFORMATION: /note= "may be T"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 994
OTHER INFORMATION: /note= "may be A"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 49..51
OTHER INFORMATION: /note= "codon might be CGG"

```

[illegible]

```

Db      1329  YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1388
QY      1182  TGTCGTTTCAATGACCAAGAAATCCAGGCCCTTGGAATTGTACTCACTAATTC 1241
Db      1389  YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYGTAACAATTCCTC 1448
QY      1242  TAACATAAGAACCCTGTGTGCACAAAGGTAAACAAGACCCTTGAGCTGT 1295
Db      1449  TATCTCTTAACTACTTGCAATGAGTAGGTAATTACAGTGAIGCCCTACATGCCGT 1502

RESULT 10
US-08-992-334-1
; Sequence 1, Application US/08992334
; Patent No. 5919678
; GENERAL INFORMATION:
; APPLICANT: Gruss, Alexandra
; APPLICANT: Maglin, Emmanuelle
; TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
; TITLE OF INVENTION: PLASMID
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christie Parker & Hale, LLP
; STREET: 350 West Colorado Boulevard, Suite 500
; City: Pasadena
; STATE: California
; COUNTRY: United States
; ZIP: 91105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/992.334
; FILING DATE: 17-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,752
; FILING DATE: 24-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR93/00248
; FILING DATE: 12-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 992/03034
; FILING DATE: 13-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Prout, D. Bruce
; REGISTRATION NUMBER: 20958
; REFERENCE/DOCKET NUMBER: C93:31779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (626) 795-9900
; TELEFAX: (626) 577-8800
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3792 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: YES
; IMMEDIATE SOURCE:
; CLONE: pg+host4
;
US-08-992-334-1

Query Match          2.4%   Score 38.4: DB 2; Length 3792:
Best Local Similarity 46.6%; Pred. NO. 0.25;
Matches 123: Conservative 0; Mismatches 141; Indels 0; Gaps 0

QY      104  AGGATTGGAGAAAAGAAATATAGCACTTTATTTCTTCTGCAAGTTATAGCAAGAAAG 163
         || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      2647  ACTGTGGATATAGACCAAGTTATGCAAGAGTTCTGTGATGCTGTAAACGGGAAATPAAAT 2706

```

Wed Feb 19 14:09:23 2003

us-09-605-042a-37.in1

```

OY 164 GGGAAAAAAGACCTTCTTATGGGGATTAACGGCGCAAGATACAAAGATTTCAGAA 223
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2707 GACCAACAAAGAAAAAGAGTATTTGCTGAAAATGAGAAATTAAGAAATTAAGCA 2766
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 224 AAGATAAATAGTACCAATGAAAGATTAAGTCTTTTCATCAAGAACCGCAAAAG 283
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2767 CTTAAAGAGCGTATGTAAGATACAGAAATGAGAGTTGAATTAAGTACACAAATGA 2826
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 284 TAAATATATATAGTATGATCTTCTCCTTTCCACAAAGATGAAGTTATGCCAGGTGG 343
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2827 TTTATGAGAGGAGGATTTATGAAATTAATAAAGCCCTGACAAAGTCAAGGGGGT 2886
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 344 CTGAGTACTGCTGGGATTCGA 367
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2887 TTTATTTTGTGTTGATGTGCGA 2910
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 11
US-08-302-752-1
Sequence 1, Application US/08302752
Patent No. 6025190
GENERAL INFORMATION:
APPLICANT: THERMOSENSIBLE PLASMID
TITLE OF INVENTION: THERMOSENSIBLE PLASMID
NUMBER OF SEQUENCES: 3
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302,752
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA: FR 9203034
APPLICATION NUMBER: 13-MAR-1992
FILING DATE: 13-MAR-1992
PRIOR APPLICATION DATA: WO FR/93/00248
APPLICATION NUMBER: NO FR/93/00248
FILING DATE: 12-MAR-1993
APPLICATION ID NO: 1
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3792 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-302-752-1
Query Match 2.4%; Score 38.4; DB 3; Length 3792;
Best Local Similarity 46.6%; Pred. No. 0.25; Indels 0; Gaps 0;
Matches 123; Conservative 0; Mismatches 141;
OY 104 AGGATGGAAGAAAGATATTAGAACTTTTATTTCTCTGAAAGTTATAGCAAGAAAG 163
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2647 AGGTGGATATAGCAAGATTAAGCAAGTTCTTCTGATCTGAAACGGGGAAATTAAT 2706
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 164 GGGAAAAAAGACCTTCTTATGGGGATTAACGGCGCAAGATACAAAGATTTCAGAA 223
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2707 GACCAACAAAGAAAAAGAGTATTTGCTGAAAATGAGAAATTAAGAAATTAAGCA 2766
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 224 AAGATAAATAGTACCAATGAAAGATTAAGTCTTTTCATCAAGAACCGCAAAAG 283
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2767 CTTAAAGAGCGTATGTAAGATACAGAAATGAGAGTTGAATTAAGTACACAAATGA 2826
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 284 TAAATATATATAGTATGATCTTCTCCTTTCCACAAAGATGAAGTTATGCCAGGTGG 343
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2827 TTTATGAGAGGAGGATTTATGAAATTAATAAAGCCCTGACAAAGTCAAGGGGGT 2886
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 344 CTGAGTACTGCTGGGATTCGA 367
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2887 TTTATTTTGTGTTGATGTGCGA 2910
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 12
US-08-992-334-2
Sequence 2, Application US/08992334
Patent No. 5919678
GENERAL INFORMATION:
APPLICANT: Gruss, Alexandra
TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE PLASMID
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESS: Christle Parker & Hale, LLP
STREET: 350 West Colorado Boulevard, Suite 500
CITY: Pasadena
STATE: California
COUNTRY: United States
ZIP: 91105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/992,334
FILING DATE: 17-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,752
FILING DATE: 24-DEC-1994
PRIOR APPLICATION DATA: PCT/FR93/00248
APPLICATION NUMBER: PCT/FR93/00248
FILING DATE: 12-MAR-1993
PRIOR APPLICATION DATA: FR 9203034
APPLICATION NUMBER: FR 9203034
FILING DATE: 13-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Prout, D Bruce
REGISTRATION NUMBER: 20958
REFERENCE/DOCKET NUMBER: C93:31779
TELEPHONE: (626) 577-9900
TELEFAX: (626) 577-8800
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5234 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
US-08-992-334-2
Query Match 2.4%; Score 38.4; DB 2; Length 5234;
Best Local Similarity 46.6%; Pred. No. 0.31; Indels 0; Gaps 0;
Matches 123; Conservative 0; Mismatches 141;
OY 104 AGGATGGAAGAAAGATATTAGAACTTTTATTTCTCTGAAAGTTATAGCAAGAAAG 163
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4089 AGGTGGATATAGCAAGATTAAGCAAGTTCTTCTGATCTGAAACGGGGAAATTAAT 4148
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 164 GGGAAAAAAGACCTTCTTATGGGGATTAACGGCGCAAGATACAAAGATTTCAGAA 223
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4149 GACCAACAAAGAAAAAGAGTATTTGCTGAAAATGAGAAATTAAGAAATTAAGCA 4208
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 224 AAGATAAATAGTACCAATGAAAGATTAAGTCTTTTCATCAAGAACCGCAAAAG 283
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2767 CTTAAAGAGCGTATGTAAGATACAGAAATGAGAGTTGAATTAAGTACACAAATGA 2826
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 284 TAAATATATATAGTATGATCTTCTCCTTTCCACAAAGATGAAGTTATGCCAGGTGG 343
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4209 CTTAAAGAGCGTATGTAAGATTAAGTATTAATAAAGCCCTGACAAAGTCAAGGGGGT 4328
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 4269 TTTATGAGAGGAGGATTTATGAAATTAATAAAGCCCTGACAAAGTCAAGGGGGT 4328
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

OY 344 CTGACTACGTCGTGGATTGGA 367
| | | | |
DB 4329 TTTTATTTTGGTTGATTGCGA 4352

RESULT 13

US-08-302-752-2
; Sequence 2, Application US/08302752
; Patent No. 6025190
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: THERMOSENSIBLE PLASMID
; NUMBER OF SEQUENCES: 3
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (ERO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,752
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9203034
; FILING DATE: 13-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO FR/93/00248
; FILING DATE: 12-MAR-1993
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5234 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-302-752-2

Query Match 2.4%; Score 38.4; DB 3; Length 5234;
Best Local Similarity 46.6%; Pred. No. 0.31;
Matches 123; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

OY 104 AGCATTCAGAAAAGATATAGCACTTTTATTTCTTCTGAAATTATAGCAAGAAG 163
| | | | |
DB 4089 AGTGTGATATAGCAAGTATGCAAGCTTCTGATGCTGAAACGGGGAAATTAAT 4148
| | | | |
OY 164 GGGAAAAAAGAACTCTTATGGGGGATTAACGGCAAGATACAAAGCTTCAGAA 223
| | | | |
DB 4149 GACAAACAAAGAAAGAGTATTTGCTGAATAAGAGAAATTAAGAAATTAAGGA 4208
| | | | |
OY 224 AAGATTAATAGTAAAGCAATGAAAGATTAATCTTTTTCATCAAGAACCGCAAG 283
| | | | |
DB 4209 CTTAAAGACCGATTTGAAGATACAGAAATGGAAGTTGAATTAAGTACAAATAGA 4268
| | | | |
OY 284 TAAATATATATAGATGTTCTCATTCTCCACAAAGATGAAGTAAATGCCAGGGTGG 343
| | | | |
DB 4269 TTTATTTGAGAGGAGGATTTATTAATTAAGAAAGCCCTGACAGAAAGTGAAGGGGT 4328
| | | | |
OY 344 CTGAGTACTGCTGGGATTTGGA 367
| | | | |
DB 4329 TTTTATTTTGGTTGATTGCGA 4352
| | | | |

RESULT 14

US-08-992-334-3
; Sequence 3, Application US/08992334
; Patent No. 5919678
; GENERAL INFORMATION:
; APPLICANT: Gruss, Alexandra
; APPLICANT: Maguin, Emmanuelle
; TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
; TITLE OF INVENTION: PLASMID
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Christie Parker & Hale, LLP
STREET: 350 West Colorado Boulevard, Suite 500
CITY: Pasadena
STATE: California
COUNTRY: United States
ZIP: 91105

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/992,334
; FILING DATE: 17-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,752
; FILING DATE: 24-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR93/00248
; FILING DATE: 12-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR FR92/03034
; FILING DATE: 13-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Prot, D. Bruce
; REGISTRATION NUMBER: 20958
; REFERENCE/DOCKET NUMBER: C93:31779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (626) 795-9900
; TELEFAX: (626) 577-8800
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6722 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
US-08-992-334-3

Query Match 2.4%; Score 38.4; DB 2; Length 6722;
Best Local Similarity 46.6%; Pred. No. 0.35;
Matches 123; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

OY 104 AGCATTCAGAAAAGATATATAGCACTTTTATTTCTTCTGAAATTATAGCAAGAAG 163
| | | | |
DB 5577 AGTGTGATATAGCAAGTATGCAAGGTCTTGATGCTGAAACGGGGAAATTAAT 5636
| | | | |
OY 164 GGGAAAAAAGAACTCTTATGGGGGATTAACGGGCAAGATACAAAGCTTCAGAA 223
| | | | |
DB 5637 GACAAACAAAGAAAGAGTATTTGCTGAATAAGAGAAATTAAGAAATTAAGGA 5696
| | | | |
OY 224 AAGATTAATAGTAAAGCAATGAAAGATTAATCTTTTTCATCAAGAACCGCAAG 283
| | | | |
DB 5697 CTTAAAGACCGATTTGAAGATACAGAAATGCAAGTTGAATTAAGTCAACAAATAGA 5756
| | | | |
OY 284 TAAATATATATAGATGTTCTCATTCTCCACAAAGATGAAGTAAATGCCAGGGTGG 343
| | | | |
DB 5757 TTTATTTGAGAGGAGGATTTATTAATTAAGAAAGCCCTGACGAAAGTGAAGGGGT 5816
| | | | |
OY 344 CTGAGTACTGCTGGGATTTGGA 367
| | | | |
DB 5817 TTTTATTTTGGTTGATTGCGA 5840
| | | | |

RESULT 15

US-08-302-752-3
; Sequence 3, Application US/08302752
; Patent No. 6025190
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: THERMOSENSIBLE PLASMID
; NUMBER OF SEQUENCES: 3

Wed Feb 19 14:09:23 2003

us-09-605-042a-37.rni

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302,752
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
PRIOR APPLICATION NUMBER: FR 9203034
FILING DATE: 13-MAR-1992
PRIOR APPLICATION DATA:
PRIOR APPLICATION NUMBER: WO FR/93/00248
FILING DATE: 12-MAR-1993
APPLICATION FOR SEQ ID NO: 3:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 6722 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-302-752-3

Query Match 2.4%; Score 38.4; DB 3; Length 6722;
Best Local Similarity 46.6%; Pred. No. 0.35; Mismatches 141; Indels 0; Gaps 0;
Matches 123; Conservative

QY 104 AGATTGAGAAAAAGATATTAGAACTTTTATTTTCTCTGAGATTATAGCAAGAAAG 163
DB 5577 AGTGTGATATAGACAGATTGCAAGGTTCTGTATGCTGAAACGGGGAATTAAT 5636
QY 164 GGAAGAAAAAAACATCTTATGGGGATTAACGGCAAGATACAAACAGTTTCAGAA 223
DB 5637 GACAAACAAAGAAAGAGTTATTTCTGAAATGAGAAATTAAGAAATTAAGAA 5696
QY 224 AAGATTAATATAGTACCAATGAAAGATTAACCTCTTTCATCAAGAACGCCAAG 283
DB 5697 CTTAAAGAGCGTATGAAAGATACAGAAATGAAAGTTGAATTAAGTACACATAGA 5756
QY 284 TAAATTAATGATTAAGTCTTCTCACTTTCCACAAGATGAAGTAATGCCAGGGTGG 343
DB 5757 TTTATGAGAGAGGATTAATGAATTAAGAGCCCTGACGAAGAGTCGAAGGGGT 5816
QY 344 CTGAGTACTGTGCTGGATTGGA 367
DB 5817 TTTTATTTTGGTTGATGTTGCGA 5840

Search completed: February 17, 2003, 16:14:37
Job time : 121.862 secs
```

Gencore version 5.1.3
2003 - 2003 CompuGen Ltd.

using sw model

(without alignments) updates/sec
12034.813 Million cell

4570410

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

N_Geneseq_101002: *
/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT: *
/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT: *
/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT: *

C	42	43.8	2.7	396	24	ABL4881
C	43	43.8	2.7	5845	24	ABL33662
C	44	43.8	2.7	10552	24	ABL70388
C	45	43.8	2.7			

RESULT 1
1547 BP.
DNA: 1547 BP.

XX
AC
AAA12451;

XX 25-JUL-2000 (first entry)
XX
DT 25-JUL-2000 (first entry)
DT

XX DNA sequence of the goat uricase gene
DE ... containing cell; urine; ... hormone; collagen;

XX kidney-specific gene; urine protein excretion; heterodimeric lectin
KW animal; antibody; heterodimeric lectin; ss. promoter; ss.

	transgen	transgen
KW	transgen	transgen
KW	transgen	transgen

XX
05 Capra sp.

XX
PN
W0200015772-A2

XX
PD
23-MAR-2000

XX 99WO-IBU1005
DE 16-SEP-1999;

XX
PR 16-SEP-1998; 98US-0100540

XX
PA (NEXT -) NEXIA BIOTECHNOLOGIES INC.XX Karatzas CN;
PT

XX
DB
WPI: 2000-271415/23

XX Nucleic acid construct containing, e.g. antibody, ...
PT ... production of protein, ...

PT

Result No.	Score	% Match	Query Length	DB ID	Description
1	1444.8	88.6	1547	21	AAAI24451
2	557.6	34.2	626	21	AAAI24459
3	266.2	16.3	620	21	AAAI24448
4	143.8	8.8	9345	21	A22939366
5	141.4	8.7	625	21	AAAI24550
6	50.6	3.1	419	22	AAI893772
7	47.6	2.9	9770	24	ABL22052
8	47.2	2.9	411	22	AAI8895
9	47.2	2.9	14920	24	ABN80146

23: /SIPS2/gcgdata/genes-2
24: Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysts of the total score distribution.

SUMMARIES

xx
ps Example 8; Flg 10B; 48pp; English.

PS Example B; Fig 10B, top.

XX The present sequence represents the partial DNA sequence of the goat

CC uronidase gene promoter. It is used to construct the nucleic acid

CC constructs of the invention. The specification describes nucleic

CC acid constructs which comprise a sequence encoding a polypeptide, a

CC promoter from a kidney-specific gene, and leader sequence that enable

CC secretion of the polypeptide, by urine-producing cells, into the urine

CC The nucleic acid constructs are used for production, in the urine of

CC transgenic animals, of polypeptides, particularly antibodies,

CC heterodimeric fertility hormones, collagen or fibrinogen, but also

CC enzymes, e.g., those that degrade an undesirable component of urine

CC such as ammonia.

CC 340 C; 332 G; 406 T; 0 other;
XX

SQ Sequence 154 / BF; 400 m; Score 1444.8; DB 21; Length 1547;

Query Match	Score	Pred. No. 0;	Indels	Gaps
88.6%		0;	0;	
99.5%		7.		

Best Local Similarity	0;	Mismatches	157
Matches 1449; Conservative			

OY	98	CCACCTGAGTGTGGAAAAAGATATTAGGAACCTTTTCTTCTCGAATTATACAA	98
Db	39	CCACCTGAGTGTGGAAAAAGATATTAGGAACCTTTTCTTCTCGAATTATACAA	
OY	158	AGAAAGGGGAAAAAAAACATTCCTTATGCGGGGATTAACGGCAAGGATACAAAGT	217
Db	99	AGAAAGGGGAAAAAAAACATTCCTTATGCGGGGATTAACGGGCAAGGATACAAAGT	158
OY	218	TCGAGAAAGATTAATAGTAAAGCAATGAAAGATTAACCTCTTTTCATCAAAACCG	277
Db	159	TCGAGAAAGATTAATAGTAAAGCAATGAAAGATTAACCTCTTTTCATCAAAACCG	218
OY	278	CAAAAGTAATATGATTAAGATCTTCTCACTTTCCACAAAGATGAAAGTAAAGCCCA	337
Db	219	CAAAAGTAATATGATTAAGATCTTCTCACTTTCCACAAAGATGAAAGTAAAGCCCA	278
OY	338	GGGTGGCTGAGTACTGTGCTGGGATTTGTAACCTACCTGTATATGATCTCTGGGGTGCT	397
Db	279	GGGTGGCTGAGTACTGTGCTGGGATTTGTAACCTACCTGTATATGATCTCTGGGGTGCT	338
OY	398	GTTTGGGAGAAACATGCGCTGAAACCTAGCTACCTCTTTCTCATGAATTTCCCTGAG	457
Db	339	GTTTGGGAGAAACATGCGCTGAAACCTAGCTACCTCTTTCTCATGAATTTCCCTGAG	398
OY	458	GAGGTAGTAGAGCCGCTGCTGATCTGCACCCGAGACCTAGGCGACAGACAAAGAGAAAGC	517
Db	399	GAGGTAGTAGAGCCGCTGCTGATCTGCACCCGAGACCTAGGCGACAGACAAAGAGAAAGC	458
OY	518	CCTCAAGAGGCAATGCTGTGATCTACTGTCAATTTTCCGTGCGACGCTGAGTTCCATG	577
Db	459	CCTCAAGAGGCAATGCTGTGATCTACTGTCAATTTTCCGTGCGACGCTGAGTTCCATG	518
OY	578	TGCGTGAATTTTCTCAATATGCGCAATTCGCATTAACGTAAGTAAAGTAAAGTAAAG	637
Db	519	TGCGTGAATTTTCTCAATATGCGCAATTCGCATTAACGTAAGTAAAGTAAAGTAAAG	578
OY	638	GCTGAACCAAGCCCTGTCAATCTCTGCCCACTGACTGGCATCAACAAACAGCATGGC	697
Db	579	GCTGAACCAAGCCCTGTCAATCTCTGCCCACTGACTGGCATCAACAAACAGCATGGC	638
OY	698	AGTGAATTCACAGATGAGACCAATGCTATATGCTTAACCTTGCAATCAATCTCTTAATG	757
Db	639	AGTGAATTCACAGATGAGACCAATGCTATATGCTTAACCTTGCAATCAATCTCTTAATG	698
OY	758	GCCATAACAATTCAGACACAGGGGTAATCTGTTTACAGATGAGAAATTAACCTGG	817
Db	699	GCCATAACAATTCAGACACAGGGGTAATCTGTTTACAGATGAGAAATTAACCTGG	758
OY	818	AAGGAAATTTCTGGCACACAAAAACGCTGACACAGATTCAGATAGACCTGACCAAG	877
Db	759	AAGGAAATTTCTGGCACACAAAAACGCTGACACAGATTCAGATAGACCTGACCAAG	818

QY	878	TCAGTCGTGATCATACAAATATATCTACTCTCTCAGAGACCTCTCTTCAATGGAATTCGA	937
Db	819	TCAGTCGTGATCATCTCAAAATATATCTACTCTCTCAGAGACCTCTCTTCAATGGAATTCGA	878
QY	938	ATTTCCTGATTCACAGAGCATCTGCTCAATGATGTCTGGAATATATCTGCTCTCTGACC	997
Db	879	ATTTCCTGATTCACAGAGCATCTGCTCAATGATGTCTGGAATATATCTGCTCTCTGACC	938
QY	998	TTCAAGCCATTCCTCAGCTCTCTCTCTGATCATTGGGACCCCAAGGGAGAGCTGCTGAATC	1057
Db	939	TTCAAGCCATTCCTCAGCTCTCTCTCTGATCATTGGGACCCCAAGGGAGAGCTGCTGAATC	998
QY	1058	TCGAGGAGATGATTTGCTTGGAAATTAAGGGGCAACAATACATCCTGGTGGGAGG	1117
Db	999	TCGAGGAGATGATTTGCTTGGAAATTAAGGGGCAACAATACATCCTGGTGGGAGG	1058
QY	1118	ATGAGACACCCCTTTCTCTCTGAGCAGCCTGCTCAGATTTCTGGCTCTGCTTGGCTTC	1177
Db	1059	ATGAGACACCCCTTTCTCTCTGAGCAGCCTGCTCAGATTTCTGGCTCTGCTTGGCTTC	1118
QY	1178	ACCTTGTGCTTTTCAATGACCAAGAAATCCCAAGGCCCTTGGAAATGTTACTCAGTTAA	1237
Db	1119	ACCTTGTGCTTTTCAATGACCAAGAAATCCCAAGGCCCTTGGAAATGTTACTCAGTTAA	1178
QY	1238	TTCTCACTAAGAAACCTCTGTGGCCAAAAGGTATTAACACAGACCTTGTAGCTGTGG	1297
Db	1179	TTCTCACTAAGAAACCTCTGTGGCCAAAAGGTATTAACACAGACCTTGTAGCTGTGG	1238
QY	1298	GCACACCTGTGACCCCATGTGCATCTTTGGGCTCTCTCCCTATTAGGAAAGAAACAA	1357
Db	1239	GCACACCTGTGACCCCATGTGCATCTTTGGGCTCTCTCCCTATTAGGAAAGAAACAA	1298
QY	1358	CAACCACTCTCAGCCTTGAAGAAAGAAACACTGTGTCAAAAAGGGAATAATTTCCACCC	1417
Db	1299	CAACCACTCTCAGCCTTGAAGAAAGAAACACTGTGTCAAAAAGGGAATAATTTCCACCC	1358
QY	1418	CCATTAAATATATTAAGAAACAGAACCCAGAGATCAATTGAGAGAGATTGCCAGTGGGG	1477
Db	1359	CCATTAAATATATTAAGAAACAGAACCCAGAGATCAATTGAGAGAGATTGCCAGTGGGG	1418
QY	1478	GACGATGTATATATATAGATATGAAGTCACTACTCTGTAAGAGATTAATTTCACTT	1537
Db	1419	GACGATGTATATATATAGATATGAAGTCACTACTCTGTAAGAGATTAATTTCACTT	1478
QY	1538	TCTGGTTTCAGTAAG 1553	
Db	1479	TCTGGTTTCAGTAAG 1494	

RESULT 2

AAA12449 standard; DNA; 626 BP.

AA
AC
AAA12449;

XX 25-JUL-2000 (first entry)

XX sequence of th

Partial DNA sequences

kidney-specific gene; urine-producing
antibody; heterodimeric fertilin

transgenic animal; gene promoter; enzyme; uromodulin gene promoter; fibroblast; uromodulin

[illegible]

Bos sp.

AA
PN
WO200015772-A2.

XX
03-MAR-2000.

PD 2070-TR01609.
XX

16-SEP-1999; 3500 1200-1300

XX

PA (NEXI-) NEXIA BIOTECHNOLOGIES INC.
 XX Karatzas CN;
 XX
 XX
 DR WPI: 2000-271415/23.
 PT Nucleic acid construct containing kidney-specific promoter, useful for
 XX recombinant production of protein, e.g. antibody, in the urine -
 PS
 XX Example 1; Fig 2; 48bp; English.

The present sequence represents the partial DNA sequence of the bovine
 CC uromodulin gene promoter. It is used to construct the nucleic acid
 CC constructs of the invention. The specification describes nucleic
 CC acid constructs which comprise a sequence encoding a polypeptide, a
 CC promoter from a kidney-specific gene, and leader sequence that enables
 CC secretion of the polypeptide, by urine-producing cells, into the urine.
 CC The nucleic acid constructs are used for production, in the urine of
 CC transgenic animals, of polypeptides, particularly antibodies,
 CC heterodimeric fertility hormones, collagen or fibrinogen, but also
 CC enzymes, e.g. those that degrade an undesirable component of urine
 CC such as ammonia.

SQ Sequence 626 BP; 184 A; 147 C; 130 G; 165 T; 0 other;

Query Match 34.2%; Score 557.6; DB 21; Length 626;
 Best local Similarity 94.1%; Pred. No. 4.3e-142;
 Matches 591; Conservative 0; Mismatches 34; Indels 3; Gaps 1;

QY 937 AATTCTTGATTCACAGAGCATCTGTCCTCAATGATCTGATTAATCTGCTGCTGAC 996
 DB 1 AATTCTTGATTCACAGAGCATCTGTCCTCAATGATCTGATTAATCTGCTGCTGAC 60
 QY 997 CTTCAGCATCTCTGAGCCCTTCCGATCACAATGGAGCCAGGGGAGCTGCTGAT 1056
 DB 61 CTTCAGGATCTCTGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 120
 QY 1057 CTCTGAGAGTGGCATTTGCTTGAATTAAGTGGCCACAAGTACATCTCTGTTGGGAC 1116
 DB 121 CCAAGAGATGGAACCTTGCTTGGAACTGAGTGGCCACAAGTATACATCCAGTGGGAC 180
 QY 1117 GATGAGCACTCTTCTCTCTGAGAGAGCTGCTCAATTTGAGCTGCTGCTGCTC 1176
 DB 181 AGTGGCACTCTTCTCTCTGAGAGAGCTGCTCAATTTGAGCTGCTGCTGCTC 237
 QY 1177 CACTTGTGCTTTCATATGACCAAGAAATCCAGGCCCTTGAATTTGTTACTAGTTA 1236
 DB 238 CACTTGTGCTTTCATATGACCAAGAAATCCAGGCCCTTGAATTTGTTACTAGTTA 297
 QY 1237 ATTTCTAATAAGAACTCTTGTGTCACAAAGATATAAAGAGCCCTTGTAGCTGTG 1296
 DB 298 ATTTCTAATAAGAACTCTGCTGTCACAAAGATATAAAGAGCCCTTGTAGCTGTG 357
 QY 1297 GGCACAGCTGTGACCCCATGTGCAATCATTTGGGGTCTTACCTATTAAGGAAAAAGACA 1356
 DB 358 GGCACAGCTGTGACCCCATGTGCAATCATTTGGGGTCTTACCTATTAAGGAAAAAGACA 417
 QY 1357 ACAACCACTCTCAGAGCTGAGAAAAAGAAACACATGCTCAAAAAGGAAAAATATTCAC 1416
 DB 418 ACAACCACTCTCAGAGCTGAGAAAAAGAAACACATGCTCAAAAAGGAAAAATATTCAC 477
 QY 1417 CCATTAATAATAATAAGAAACGAACACAGAGATCATTTGGAGAGAGATTTGCCAGTGG 1476
 DB 478 CCATTAATAATAATAAGAAACGAACACAGAGATCATTTGGAGAGAGATTTGCCAGTGG 537
 QY 1477 GGACAGATGATATATATATATGATATGAAAGTCACTACTTGTAAAAAGATTAATTTACT 1536
 DB 538 GGACAGATGATATATATATATGATATGATATGATATGATATGATATGATATGATAT 597
 QY 1537 TTTCTGTTTCAAGTAAAGGCTATCTGAG 1564
 DB 598 TTTCTGTTTCAAGTAAAGGCTATCTGAG 625

RESULT 3
 ID AAA12448
 AC AAA12448 standard; DNA; 620 BP.
 XX
 AC AAA12448;
 XX
 DT 25-JUL-2000 (first entry)
 XX
 DE Partial DNA sequence of the human uromodulin gene promoter.
 XX
 KW kidney-specific gene; urine-producing cell; urine;
 KW transgenic animal; antibody; heterodimeric fertility hormone; collagen;
 KW fibrinogen; enzyme; uromodulin gene promoter; ss.
 OS Homo sapiens.
 PN WO200015772-A2.
 XX
 PD 23-MAR-2000.
 XX
 PF 16-SEP-1999; 99WO-IB01609.
 XX
 PR 16-SEP-1998; 98US-0100540.
 XX
 PA (NEXI-) NEXIA BIOTECHNOLOGIES INC.
 XX Karatzas CN;
 XX
 DR WPI: 2000-271415/23.

Nucleic acid construct containing kidney-specific promoter, useful for
 PT recombinant production of protein, e.g. antibody, in the urine -
 PS
 XX Disclosure; Fig 1; 48bp; English.

The present sequence represents the partial DNA sequence of the human
 CC uromodulin gene promoter. It is used to construct the nucleic acid
 CC constructs of the invention. The specification describes nucleic
 CC acid constructs which comprise a sequence encoding a polypeptide, a
 CC promoter from a kidney-specific gene, and leader sequence that enables
 CC secretion of the polypeptide, by urine-producing cells, into the urine.
 CC The nucleic acid constructs are used for production, in the urine of
 CC transgenic animals, of polypeptides, particularly antibodies,
 CC heterodimeric fertility hormones, collagen or fibrinogen, but also
 CC enzymes, e.g. those that degrade an undesirable component of urine
 CC such as ammonia.

SQ Sequence 620 BP; 164 A; 155 C; 138 G; 163 T; 0 other;

Query Match 16.3%; Score 266.2; DB 21; Length 620;
 Best local Similarity 72.8%; Pred. No. 1.8e-62;
 Matches 461; Conservative 0; Mismatches 143; Indels 29; Gaps 8;

QY 949 CACAGACATCTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1008
 DB 1 CAGAGTGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 60
 QY 1009 TCAAGCTCTTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1068
 DB 61 TCAAGCTCTTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 113
 QY 1069 CATTTGCTTGAATTAAGTGGCC---ACAAGTACATCTCTGATGATGATGATGATG 1125
 DB 114 GATTTTCAAGAGGGTGGCTCTCAAGACAGGTGACCTTCCATCTGAGGACATGATAT 1173
 QY 1126 CCTTTTCTCTGAGAGCCGCTTCAAGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1177
 DB 174 -CCTTTTCTCTGAGAGCCGCTTCAAGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 232
 QY 1178 ACTTTGCTTCTTCAATGACCAAGAA-ATTCAGAGCCCTTGAATTTGTTACTCAAGTTA 1236
 DB 233 ACCCTGACCTTGTAGTACCAAGAAAGACCAAGCAGCCTTGGACAGCTTTCAGAGTTA 292

Wed Feb 19 14:09:22 2003

us-09-605-042a-37.ing

```

PR 09-JUL-1999; 99US-0142925.
PA (UNYNT) UNTV NEW YORK STATE.
PI Wu X, Sun T;
XX WPI: 2000-387816/33.
DR New kidney-specific promoter useful for production of transgenic
PT animals as urinary bioeffectors, is operably linked to a heterologous
PT gene
PS Claim 3; Fig 7a-7h: 55pp; English.
XX
XX New methods to produce heterologous recombinant proteins in urine
XX require the use of a DNA molecule which is a kidney-specific
XX promoter, such as the uromodulin promoter operably linked to a
XX heterologous gene encoding a biologically active protein. The
XX uromodulin promoter expresses the heterologous active protein in the
XX kidneys to produce a recombinant protein produced may be useful for treating
XX human diseases. The major advantages of using this urine-based system
XX after milk-based systems are the ability to harvest the product soon
XX after birth and throughout the life of the animal irrespective of sex
XX or reproductive status, and the ease of product purification from
XX urine. In addition, livestock urine is a proven, currently utilized
XX source of pharmaceuticals.
SQ
Sequence 9345 BP: 2622 A; 2072 C; 2084 G; 2565 T; 2 other:
Query Match 8.8%; Score 143.8; DB 21; Length 9345;
Best Local Similarity 65.7%; Pred. No. 1.8e-28; Indels 7; Gaps 5;
Matches 286; Conservative 0; Mismatches 142;
QY 1163 CCTGCTGGCTGCACCTTTGCTTTCATGACCAAGAAA-TCGAGGCCCTTGAA 1221
DB 6737 CTGGGCTTAGCTCTACCTTGTCTCCAGAGTCCCAAGAAAAGCCCAACTTGCA 6796
QY 1222 TTGTTTACTGATTAATTTCTTAACAAACCTCTGTTGCCAAAGGTATAACGA 1281
DB 6797 CTGTTTCCAGTTAATGCTAATCTGAGAAATGCTTGCCCAAAAGT-GAAAACGA 6855
QY 1282 GCCCTTGTAGCTGTGGGACAGCTGTGACCCCATGTCATCATTTGGGGTCT-CTACCT 1340
DB 6856 GACCTTGTATTTCCAGGACAGAGTGTGACCCCAATGTCAATCTTTGTCTTAACCTCC 6915
QY 1341 ATTAGGAAAGAACACACACACCACTCAGAGCTGAGAAAGGAAAACACAGTGTGCAAAAG 1400
DB 6916 AGGGGAAAACCTAACAAACAGACTCATAGTGTGAGAAAGTGTATTTGCAAAAG 6975
QY 1401 GGAATAATATCCACCCCATTAATTAATTAAGAAACGACAGAGATCTTGGAAG 7035
DB 6976 GGAAGGAAAGTTCTACCCCAAGAAACAATCTCAGAGGCAAGAGATTAATCTG 7095
QY 1461 AGAGATTGCCAGTGGGGGACAGATGTATATATATAGATTAAGAAAGTCACTTGTAA 1520
DB 7036 AGGAGAGGAGCCAGCAGAGGCAAGCAATATATATATGACAGGCACTTACTGTGA 7095
QY 1521 AGGATTAATCTACCTTCTTGTTTCAAGTAAAGCTATCTGCACTCTCCTCTAG 1580
DB 7096 TGGACCAAGTCT---GTCTGTGGTTCAGGTAAAGCTGTATGAATCTGTAC-CCCAAT 7151
QY 1581 CCACTTCTCCCACT 1595
DB 7152 CCACTTCTCTCTAT 7166

```

RESULT 4
 AA293966
 ID AA293966 standard; DNA: 9345 BP.
 AC AA293966;
 DT 29-AUG-2000 (first entry)
 XX
 DE Mouse uromodulin promoter.
 XX
 KW Uromodulin; promoter; kidney; urine; heterologous gene; treatment;
 XX therapy; gene expression; pharmaceutical; mouse; ds.
 OS Mus musculus.
 XX
 FH Location/Qualifiers
 FT 4..5
 FT /*tag= a
 FT /label= Polycloning site of pBS
 FT 7095..7114
 FT /*tag= b
 FT /label= Exon 1
 FT /cons-splice= (5'site:NO, 3'site:YES)
 FT 7115..7892
 FT /*tag= c
 FT /label= Intron 1
 FT 7893..8073
 FT /*tag= d
 FT /label= Exon 2
 FT /cons-splice= (5'site:NO, 3'site:YES)
 FT 8074..8818
 FT /*tag= e
 FT /label= Intron 2
 FT 8819..9345
 FT /*tag= f
 FT /note= "APN/KPN Junction"
 FT /label= Exon 3
 FT exon
 FT misc-feature
 FT 8273..8274
 FT /*tag= g
 FT /label= Intron 3
 FT 8275..8276
 FT /*tag= h
 FT /label= Exon 4
 FT 8277..8278
 FT /*tag= i
 FT /label= Intron 4
 FT 8279..8280
 FT /*tag= j
 FT /label= Exon 5
 FT 8281..8282
 FT /*tag= k
 FT /label= Intron 5
 FT 8283..8284
 FT /*tag= l
 FT /label= Exon 6
 FT 8285..8286
 FT /*tag= m
 FT /label= Intron 6
 FT 8287..8288
 FT /*tag= n
 FT /label= Exon 7
 FT 8289..8290
 FT /*tag= o
 FT /label= Intron 7
 FT 8291..8292
 FT /*tag= p
 FT /label= Exon 8
 FT 8293..8294
 FT /*tag= q
 FT /label= Intron 8
 FT 8295..8296
 FT /*tag= r
 FT /label= Exon 9
 FT 8297..8298
 FT /*tag= s
 FT /label= Intron 9
 FT 8299..8300
 FT /*tag= t
 FT /label= Exon 10
 FT 8301..8302
 FT /*tag= u
 FT /label= Intron 10
 FT 8303..8304
 FT /*tag= v
 FT /label= Exon 11
 FT 8305..8306
 FT /*tag= w
 FT /label= Intron 11
 FT 8307..8308
 FT /*tag= x
 FT /label= Exon 12
 FT 8309..8310
 FT /*tag= y
 FT /label= Intron 12
 FT 8311..8312
 FT /*tag= z
 FT /label= Exon 13
 FT 8313..8314
 FT /*tag= AA
 FT /label= Intron 13
 FT 8315..8316
 FT /*tag= AB
 FT /label= Exon 14
 FT 8317..8318
 FT /*tag= AC
 FT /label= Intron 14
 FT 8319..8320
 FT /*tag= AD
 FT /label= Exon 15
 FT 8321..8322
 FT /*tag= AE
 FT /label= Intron 15
 FT 8323..8324
 FT /*tag= AF
 FT /label= Exon 16
 FT 8325..8326
 FT /*tag= AG
 FT /label= Intron 16
 FT 8327..8328
 FT /*tag= AH
 FT /label= Exon 17
 FT 8329..8330
 FT /*tag= AI
 FT /label= Intron 17
 FT 8331..8332
 FT /*tag= AJ
 FT /label= Exon 18
 FT 8333..8334
 FT /*tag= AK
 FT /label= Intron 18
 FT 8335..8336
 FT /*tag= AL
 FT /label= Exon 19
 FT 8337..8338
 FT /*tag= AM
 FT /label= Intron 19
 FT 8339..8340
 FT /*tag= AN
 FT /label= Exon 20
 FT 8341..8342
 FT /*tag= AO
 FT /label= Intron 20
 FT 8343..8344
 FT /*tag= AP
 FT /label= Exon 21
 FT 8345..8346
 FT /*tag= AQ
 FT /label= Intron 21
 FT 8347..8348
 FT /*tag= AR
 FT /label= Exon 22
 FT 8349..8350
 FT /*tag= AS
 FT /label= Intron 22
 FT 8351..8352
 FT /*tag= AT
 FT /label= Exon 23
 FT 8353..8354
 FT /*tag= AU
 FT /label= Intron 23
 FT 8355..8356
 FT /*tag= AV
 FT /label= Exon 24
 FT 8357..8358
 FT /*tag= AW
 FT /label= Intron 24
 FT 8359..8360
 FT /*tag= AX
 FT /label= Exon 25
 FT 8361..8362
 FT /*tag= AY
 FT /label= Intron 25
 FT 8363..8364
 FT /*tag= AZ
 FT /label= Exon 26
 FT 8365..8366
 FT /*tag= BA
 FT /label= Intron 26
 FT 8367..8368
 FT /*tag= BB
 FT /label= Exon 27
 FT 8369..8370
 FT /*tag= BC
 FT /label= Intron 27
 FT 8371..8372
 FT /*tag= BD
 FT /label= Exon 28
 FT 8373..8374
 FT /*tag= BE
 FT /label= Intron 28
 FT 8375..8376
 FT /*tag= BF
 FT /label= Exon 29
 FT 8377..8378
 FT /*tag= BG
 FT /label= Intron 29
 FT 8379..8380
 FT /*tag= BH
 FT /label= Exon 30
 FT 8381..8382
 FT /*tag= BI
 FT /label= Intron 30
 FT 8383..8384
 FT /*tag= BJ
 FT /label= Exon 31
 FT 8385..8386
 FT /*tag= BK
 FT /label= Intron 31
 FT 8387..8388
 FT /*tag= BL
 FT /label= Exon 32
 FT 8389..8390
 FT /*tag= BM
 FT /label= Intron 32
 FT 8391..8392
 FT /*tag= BN
 FT /label= Exon 33
 FT 8393..8394
 FT /*tag= BO
 FT /label= Intron 33
 FT 8395..8396
 FT /*tag= BP
 FT /label= Exon 34
 FT 8397..8398
 FT /*tag= BQ
 FT /label= Intron 34
 FT 8399..8400
 FT /*tag= BR
 FT /label= Exon 35
 FT 8401..8402
 FT /*tag= BS
 FT /label= Intron 35
 FT 8403..8404
 FT /*tag= BT
 FT /label= Exon 36
 FT 8405..8406
 FT /*tag= BU
 FT /label= Intron 36
 FT 8407..8408
 FT /*tag= BV
 FT /label= Exon 37
 FT 8409..8410
 FT /*tag= BW
 FT /label= Intron 37
 FT 8411..8412
 FT /*tag= BX
 FT /label= Exon 38
 FT 8413..8414
 FT /*tag= BY
 FT /label= Intron 38
 FT 8415..8416
 FT /*tag= BZ
 FT /label= Exon 39
 FT 8417..8418
 FT /*tag= CA
 FT /label= Intron 39
 FT 8419..8420
 FT /*tag= CB
 FT /label= Exon 40
 FT 8421..8422
 FT /*tag= CC
 FT /label= Intron 40
 FT 8423..8424
 FT /*tag= CD
 FT /label= Exon 41
 FT 8425..8426
 FT /*tag= CE
 FT /label= Intron 41
 FT 8427..8428
 FT /*tag= CF
 FT /label= Exon 42
 FT 8429..8430
 FT /*tag= CG
 FT /label= Intron 42
 FT 8431..8432
 FT /*tag= CH
 FT /label= Exon 43
 FT 8433..8434
 FT /*tag= CI
 FT /label= Intron 43
 FT 8435..8436
 FT /*tag= CJ
 FT /label= Exon 44
 FT 8437..8438
 FT /*tag= CK
 FT /label= Intron 44
 FT 8439..8440
 FT /*tag= CL
 FT /label= Exon 45
 FT 8441..8442
 FT /*tag= CM
 FT /label= Intron 45
 FT 8443..8444
 FT /*tag= CN
 FT /label= Exon 46
 FT 8445..8446
 FT /*tag= CO
 FT /label= Intron 46
 FT 8447..8448
 FT /*tag= CP
 FT /label= Exon 47
 FT 8449..8450
 FT /*tag= CQ
 FT /label= Intron 47
 FT 8451..8452
 FT /*tag= CR
 FT /label= Exon 48
 FT 8453..8454
 FT /*tag= CS
 FT /label= Intron 48
 FT 8455..8456
 FT /*tag= CT
 FT /label= Exon 49
 FT 8457..8458
 FT /*tag= CU
 FT /label= Intron 49
 FT 8459..8460
 FT /*tag= CV
 FT /label= Exon 50
 FT 8461..8462
 FT /*tag= CW
 FT /label= Intron 50
 FT 8463..8464
 FT /*tag= CX
 FT /label= Exon 51
 FT 8465..8466
 FT /*tag= CY
 FT /label= Intron 51
 FT 8467..8468
 FT /*tag= CZ
 FT /label= Exon 52
 FT 8469..8470
 FT /*tag= DA
 FT /label= Intron 52
 FT 8471..8472
 FT /*tag= DB
 FT /label= Exon 53
 FT 8473..8474
 FT /*tag= DC
 FT /label= Intron 53
 FT 8475..8476
 FT /*tag= DD
 FT /label= Exon 54
 FT 8477..8478
 FT /*tag= DE
 FT /label= Intron 54
 FT 8479..8480
 FT /*tag= DF
 FT /label= Exon 55
 FT 8481..8482
 FT /*tag= DG
 FT /label= Intron 55
 FT 8483..8484
 FT /*tag= DH
 FT /label= Exon 56
 FT 8485..8486
 FT /*tag= DI
 FT /label= Intron 56
 FT 8487..8488
 FT /*tag= DJ
 FT /label= Exon 57
 FT 8489..8490
 FT /*tag= DK
 FT /label= Intron 57
 FT 8491..8492
 FT /*tag= DL
 FT /label= Exon 58
 FT 8493..8494
 FT /*tag= DM
 FT /label= Intron 58
 FT 8495..8496
 FT /*tag= DN
 FT /label= Exon 59
 FT 8497..8498
 FT /*tag= DO
 FT /label= Intron 59
 FT 8499..8500
 FT /*tag= DP
 FT /label= Exon 60
 FT 8501..8502
 FT /*tag= DQ
 FT /label= Intron 60
 FT 8503..8504
 FT /*tag= DR
 FT /label= Exon 61
 FT 8505..8506
 FT /*tag= DS
 FT /label= Intron 61
 FT 8507..8508
 FT /*tag= DT
 FT /label= Exon 62
 FT 8509..8510
 FT /*tag= DU
 FT /label= Intron 62
 FT 8511..8512
 FT /*tag= DV
 FT /label= Exon 63
 FT 8513..8514
 FT /*tag= DW
 FT /label= Intron 63
 FT 8515..8516
 FT /*tag= DX
 FT /label= Exon 64
 FT 8517..8518
 FT /*tag= DY
 FT /label= Intron 64
 FT 8519..8520
 FT /*tag= DZ
 FT /label= Exon 65
 FT 8521..8522
 FT /*tag= EA
 FT /label= Intron 65
 FT 8523..8524
 FT /*tag= EB
 FT /label= Exon 66
 FT 8525..8526
 FT /*tag= EC
 FT /label= Intron 66
 FT 8527..8528
 FT /*tag= ED
 FT /label= Exon 67
 FT 8529..8530
 FT /*tag= EE
 FT /label= Intron 67
 FT 8531..8532
 FT /*tag= EF
 FT /label= Exon 68
 FT 8533..8534
 FT /*tag= EG
 FT /label= Intron 68
 FT 8535..8536
 FT /*tag= EH
 FT /label= Exon 69
 FT 8537..8538
 FT /*tag= EI
 FT /label= Intron 69
 FT 8539..8540
 FT /*tag= EJ
 FT /label= Exon 70
 FT 8541..8542
 FT /*tag= EK
 FT /label= Intron 70
 FT 8543..8544
 FT /*tag= EL
 FT /label= Exon 71
 FT 8545..8546
 FT /*tag= EM
 FT /label= Intron 71
 FT 8547..8548
 FT /*tag= EN
 FT /label= Exon 72
 FT 8549..8550
 FT /*tag= EO
 FT /label= Intron 72
 FT 8551..8552
 FT /*tag= EP
 FT /label= Exon 73
 FT 8553..8554
 FT /*tag= EQ
 FT /label= Intron 73
 FT 8555..8556
 FT /*tag= ER
 FT /label= Exon 74
 FT 8557..8558
 FT /*tag= ES
 FT /label= Intron 74
 FT 8559..8560
 FT /*tag= ET
 FT /label= Exon 75
 FT 8561..8562
 FT /*tag= EU
 FT /label= Intron 75
 FT 8563..8564
 FT /*tag= EV
 FT /label= Exon 76
 FT 8565..8566
 FT /*tag= EW
 FT /label= Intron 76
 FT 8567..8568
 FT /*tag= EX
 FT /label= Exon 77
 FT 8569..8570
 FT /*tag= EY
 FT /label= Intron 77
 FT 8571..8572
 FT /*tag= EZ
 FT /label= Exon 78
 FT 8573..8574
 FT /*tag= FA
 FT /label= Intron 78
 FT 8575..8576
 FT /*tag= FB
 FT /label= Exon 79
 FT 8577..8578
 FT /*tag= FC
 FT /label= Intron 79
 FT 8579..8580
 FT /*tag= FD
 FT /label= Exon 80
 FT 8581..8582
 FT /*tag= FE
 FT /label= Intron 80
 FT 8583..8584
 FT /*tag= FF
 FT /label= Exon 81
 FT 8585..8586
 FT /*tag= FG
 FT /label= Intron 81
 FT 8587..8588
 FT /*tag= FH
 FT /label= Exon 82
 FT 8589..8590
 FT /*tag= FI
 FT /label= Intron 82
 FT 8591..8592
 FT /*tag= FJ
 FT /label= Exon 83
 FT 8593..8594
 FT /*tag= FK
 FT /label= Intron 83
 FT 8595..8596
 FT /*tag= FL
 FT /label= Exon 84
 FT 8597..8598
 FT /*tag= FM
 FT /label= Intron 84
 FT 8599..8600
 FT /*tag= FN
 FT /label= Exon 85
 FT 8601..8602
 FT /*tag= FO
 FT /label= Intron 85
 FT 8603..8604
 FT /*tag= FP
 FT /label= Exon 86
 FT 8605..8606
 FT /*tag= FQ
 FT /label= Intron 86
 FT 8607..8608
 FT /*tag= FR
 FT /label= Exon 87
 FT 8609..8610
 FT /*tag= FS
 FT /label= Intron 87
 FT 8611..8612
 FT /*tag= FT
 FT /label= Exon 88
 FT 8613..8614
 FT /*tag= FU
 FT /label= Intron 88
 FT 8615..8616
 FT /*tag= FV
 FT /label= Exon 89
 FT 8617..8618
 FT /*tag= FW
 FT /label= Intron 89
 FT 8619..8620
 FT /*tag= FX
 FT /label= Exon 90
 FT 8621..8622
 FT /*tag= FY
 FT /label= Intron 90
 FT 8623..8624
 FT /*tag= FZ
 FT /label= Exon 91
 FT 8625..8626
 FT /*tag= GA
 FT /label= Intron 91
 FT 8627..8628
 FT /*tag= GB
 FT /label= Exon 92
 FT 8629..8630
 FT /*tag= GC
 FT /label= Intron 92
 FT 8631..8632
 FT /*tag= GD
 FT /label= Exon 93
 FT 8633..8634
 FT /*tag= GE
 FT /label= Intron 93
 FT 8635..8636
 FT /*tag= GF
 FT /label= Exon 94
 FT 8637..8638
 FT /*tag= GG
 FT /label= Intron 94
 FT 8639..8640
 FT /*tag= GH
 FT /label= Exon 95
 FT 8641..8642
 FT /*tag= GI
 FT /label= Intron 95
 FT 8643..8644
 FT /*tag= GJ
 FT /label= Exon 96
 FT 8645..8646
 FT /*tag= GK
 FT /label= Intron 96
 FT 8647..8648
 FT /*tag= GL
 FT /label= Exon 97
 FT 8649..8650
 FT /*tag= GM
 FT /label= Intron 97
 FT 8651..8652
 FT /*tag= GN
 FT /label= Exon 98
 FT 8653..8654
 FT /*tag= GO
 FT /label= Intron 98
 FT 8655..8656
 FT /*tag= GP
 FT /label= Exon 99
 FT 8657..8658
 FT /*tag= GQ
 FT /label= Intron 99
 FT 8659..8660
 FT /*tag= GR
 FT /label= Exon 100
 FT 8661..8662
 FT /*tag= GS
 FT /label= Intron 100
 FT 8663..8664
 FT /*tag= GT
 FT /label= Exon 101
 FT 8665..8666
 FT /*tag= GU
 FT /label= Intron 101
 FT 8667..8668
 FT /*tag= GV
 FT /label= Exon 102
 FT 8669..8670
 FT /*tag= GW
 FT /label= Intron 102
 FT 8671..8672
 FT /*tag= GX
 FT /label= Exon 103
 FT 8673..8674
 FT /*tag= GY
 FT /label= Intron 103
 FT 8675..8676
 FT /*tag= GZ
 FT /label= Exon 104
 FT 8677..8678
 FT /*tag= HA
 FT /label= Intron 104
 FT 8679..8680
 FT /*tag= HB
 FT /label= Exon 105
 FT 8681..8682
 FT /*tag= HC
 FT /label= Intron 105
 FT 8683..8684
 FT /*tag= HD
 FT /label= Exon 106
 FT 8685..8686
 FT /*tag= HE
 FT /label= Intron 106
 FT 8687..8688
 FT /*tag= HF
 FT /label= Exon 107
 FT 8689..8690
 FT /*tag= HG
 FT /label= Intron 107
 FT 8691..8692
 FT /*tag= HH
 FT /label= Exon 108
 FT 8693..8694
 FT /*tag= HI
 FT /label= Intron 108
 FT 8695..8696
 FT /*tag= HJ
 FT /label= Exon 109
 FT 8697..8698
 FT /*tag= HK
 FT /label= Intron 109
 FT 8699..8700
 FT /*tag= HL
 FT /label= Exon 110
 FT 8701..8702
 FT /*tag= HM
 FT /label= Intron 110
 FT 8703..8704
 FT /*tag= HN
 FT /label= Exon 111
 FT 8705..8706
 FT /*tag= HO
 FT /label= Intron 111
 FT 8707..8708
 FT /*tag= HP
 FT /label= Exon 112
 FT 8709..8710
 FT /*tag= HQ
 FT /label= Intron 112
 FT 8711..8712
 FT /*tag= HR
 FT /label= Exon 113
 FT 8713..8714
 FT /*tag= HS
 FT /label= Intron 113
 FT 8715..8716
 FT /*tag= HT
 FT /label= Exon 114
 FT 8717..8718
 FT /*tag= HU
 FT /label= Intron 114
 FT 8719..8720
 FT /*tag= HV
 FT /label= Exon 115
 FT 8721..8722
 FT /*tag= HW
 FT /label= Intron 115
 FT 8723..8724
 FT /*tag= HX
 FT /label= Exon 116
 FT 8725..8726
 FT /*tag= HY
 FT /label= Intron 116
 FT 8727..8728
 FT /*tag= HZ
 FT /label= Exon 117
 FT 8729..8730
 FT /*tag= IA
 FT /label= Intron 117
 FT 8731..8732
 FT /*tag= IB
 FT /label= Exon 118
 FT 8733..8734
 FT /*tag= IC
 FT /label= Intron 118
 FT 8735..8736
 FT /*tag= ID
 FT /label= Exon 119
 FT 8737..8738
 FT /*tag= IE
 FT /label= Intron 119
 FT 8739..8740
 FT /*tag= IF
 FT /label= Exon 120
 FT 8741..8742
 FT /*tag= IG
 FT /label= Intron 120
 FT 8743..8744
 FT /*tag= IH
 FT /label= Exon 121
 FT 8745..8746
 FT /*tag= II
 FT /label= Intron 121
 FT 8747..8748
 FT /*tag= IJ
 FT /label= Exon 122
 FT 8749..8750
 FT /*tag= IK
 FT /label= Intron 122
 FT 8751..8752
 FT /*tag= IL
 FT /label= Exon 123
 FT 8753..8754
 FT /*tag= IM
 FT /label= Intron 123
 FT 8755..8756
 FT /*tag= IN
 FT /label= Exon 124
 FT 8757..8758
 FT /*tag= IO
 FT /label= Intron 124
 FT 8759..8760
 FT /*tag= IP
 FT /label= Exon 125
 FT 8761..8762
 FT /*tag= IQ
 FT /label= Intron 125
 FT 8763..8764
 FT /*tag= IR
 FT /label= Exon 126
 FT 8765..8766
 FT /*tag= IS
 FT /label= Intron 126
 FT 8767..8768
 FT /*tag= IT
 FT /label= Exon 127
 FT 8769..8770
 FT /*tag= IU
 FT /label= Intron 127
 FT 8771..8772
 FT /*tag= IV
 FT /label= Exon 128
 FT 8773..8774
 FT /*tag= IW
 FT /label= Intron 128
 FT 8775..8776
 FT /*tag= IX
 FT /label= Exon 129
 FT 8777..8778
 FT /*tag= IY
 FT /label= Intron 129
 FT 8779..8780
 FT /*tag= IZ
 FT /label= Exon 130
 FT 8781..8782
 FT /*tag= JA
 FT /label= Intron 130
 FT 8783..8784
 FT /*tag= JB
 FT /label= Exon 131
 FT 8785..8786
 FT /*tag= JC
 FT /label= Intron 131
 FT 8787..8788
 FT /*tag= JD
 FT /label= Exon 132
 FT 8789..8790
 FT /*tag= JE
 FT /label= Intron 132
 FT 8791..8792
 FT /*tag= JF
 FT /label= Exon 133
 FT 8793..8794
 FT /*tag= JG
 FT /label= Intron 133
 FT 8795..8796
 FT /*tag= JH
 FT /label= Exon 134
 FT 8797..8798
 FT /*tag= JI
 FT /label= Intron 134
 FT 8799..8800
 FT /*tag= JJ
 FT /label= Exon 135
 FT 8801..8802
 FT /*tag= JK
 FT /label= Intron 135
 FT 8803..8804
 FT /*tag= JL
 FT /label= Exon 136
 FT 8805..8806
 FT /*tag= JM
 FT /label= Intron 136
 FT 8807..8808
 FT /*tag= JN
 FT /label= Exon 137
 FT 8809..8810
 FT /*tag= JO
 FT /label= Intron 137
 FT 8811..8812
 FT /*tag= JP
 FT /label= Exon 138
 FT 8813..8814
 FT /*tag= JQ
 FT /label= Intron 138
 FT 8815..8816
 FT /*tag= JR
 FT /label= Exon 139
 FT 8817..8818
 FT /*tag= JS
 FT /label= Intron 139
 FT 8819..8820
 FT /*tag= JT
 FT /label= Exon 140
 FT 8821..8822
 FT /*tag= JU
 FT /label= Intron 140
 FT 8823..8824
 FT /*tag= JV
 FT /label= Exon 141
 FT 8825..8826
 FT /*tag= JW
 FT /label= Intron 141
 FT 8827..8828
 FT /*tag= JX
 FT /label= Exon 142
 FT 8829..8830
 FT /*tag= JY
 FT /label= Intron 142
 FT 8831..8832
 FT /*tag= JZ
 FT /label= Exon 143
 FT 8833..8834
 FT /*tag= KA
 FT /label= Intron 143
 FT 8835..8836
 FT /*tag= KB
 FT /label= Exon 144
 FT 8837..8838
 FT /*tag= KC
 FT /label= Intron 144
 FT 8839..8840
 FT /*tag= KD
 FT /label= Exon 145
 FT 8841..8842
 FT /*tag= KE
 FT /label= Intron 145
 FT 8843..8844
 FT /*tag= KF
 FT /label= Exon 146
 FT 8845..8846
 FT /*tag= KG
 FT /label= Intron 146
 FT 8847..8848

XX Sequence 17183 BP; 5157 A; 149 C; 320 G; 60 T
SQ

Query Match Best Local Similarity 55.6%; Score 46.9; DB 24; Length 17183;
Matches 90; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

OY 142 TCCTGACGTATTAGCAAGAAGGGGAATTAATAAACCATTCTTATGCGGCATAAACGCCG 201
Db 7742 TTTCATAATTTAATTAATTAATCCTAAAAATAAAACAATTTTATTTAAAAAATACC 7683

OY 202 AAAGCATCCAAACTTGCAAGAAAGTAATTAAGTAAAGCAAAGTAATCTCTT 261
Db 7682 CTTPAATTAATAATCTCAACAAAAATCAAAAAAAAAAACAAAACTAAACCTTAA 7623

OY 262 TTCATCAAAAGAACCGCAAGAGTAATATATGATAGTAGTTT 303
Db 7622 TAICAAAAAACCAACAAAAATTAATTTTTAAAAAACATT 7581

RESULT 11
ABL93053/C standard; cDNA: 321 BP.
ID ABL93053 standard; cDNA: 321 BP.

XX ABL93053:
AC XX
DT 10-JUN-2002 (first entry)
DE Rat metastatic tumour cell related CDNA SEQ ID NO 107.
XX Metastatic: tumour; breast cancer; pancreatic.cancer; cytostatic; ss.
XX Rattus sp.
OS WO200208456-A2.
PN 31-JAN-2002.
PD 24-JUL-2001; 2001WO-EPO8517.
PF 25-JUL-2000; 2000DE-1036126.
PR (GESL) FORSCHUNGSENZENTRUM KARLSRUHE GMBH.
PA Von Stein O, Nestl A, Hofmann M, Sleeman J, Herrlich P;
PI WPI: 2002-172012/22.
DR Identifying metastatic tumor cells and screening for antitumor agents,
PT comprises hybridization to specific CDNA sequences -
PJ Disclosure: Page 72; 140pp; German.

PS The invention relates to identifying metastatic tumour cells comprising
XX testing them for hybridisation to at least one CDNA sequence
XX (ABLP2947-ABLP3235) or the derived complete gene, functional fragments,
XX homologues or alleles. The CDNA sequences are useful for diagnosis,
XX particularly in breast and pancreatic tumours, useful for therapy and used for
XX potential as target agents for treatment of cancer, e.g. antibodies
XX identifying potentially suitable targets for treatment of cancer, allowing
XX against proteins encoded by the CDNA sequences or antisense sequences. Its
XX functional analysis, including study of metastasis-specific regulatory mechanisms of
XX metastasis

XX Sequence 321 BP; 55 A; 59 C; 30 G; 171 T; 6 other;

Query Match 2.98; Score 46.6; DB 24; Length 321;
Best Local Similarity 52.38; 1; Mismatches 0; Gaps 0;
Matches 101; Conservative

OY 103 TAGAGTATGAGAAAAACATATTAGCACTTTATTTTCCTTCGTGCAGTTATAGCAAAGAAA 162
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 200 TAGTTTGCGGAAGGGGTTTCTCCTTTGTGTTTGGAAGAAAVAAAGAAA 141
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 163 GGGGAAAAAAAAAACCTTTCTTATGGGCGATPAAACGGCCAAAGCATACAAAGCTTCACA 222
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 140 AAAAAAAAAAAAAAGGATCCAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 81
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 223 AAAGATTAAATGTATGCAATGAAAGATACTTCCTTTTCATCAAGACCGCGAAA 282
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 80 AA 21
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 283 GTAAATATGATA 295
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 20 AAAAAAAAAAAAAA 8
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 12
AB067061/C
ID AB067061 standard; DNA; 9095 BP.

AB067061:
28-AUG-2002 (first entry)

Human angiogenesis associated polynucleotide SEQ ID NO 91.
Human; angiogenesis; methylation; eye disease; glaucoma; tumour;
inflammation; rheumatoid arthritis; diabetic retinopathy; antiucercids;
macular degeneration; inflammatory bowel disease; Crohn's disease;
antirheumatic; antithyritic; antidiabetic; antiporiatic;
antiarteriosclerotic; ds.
Homo sapiens.
WO200246454-A2.
13-JUN-2002.
06-DEC-2001; 2001WO-EPI4320.
06-DEC-2000; 2000DE-1061338.
(EPIC-) EPIGENOMICS AG.
Schacht O;
WPI; 2002-500450/53.
New nucleic acid fragments from chemically treated
angiogenesis-associated genes, useful for determining methylation
status, e.g. in diagnosis or treatment of cancer -
Claim 1; SEQ ID NO 91; 41pp + Sequence Listing; German.
The invention relates to a nucleic acid (I) comprising a segment of 18.
bases of chemically pretreated DNA of angiogenesis-associated genes (II), also
having sequences (AB06971-AB067178) or their complements status and/or
related oligonucleotide polymorphisms, in angiogenesis-related diseases,
single-nucleotide treatment of eye diseases, inflammation, rheumatoid arthritis,
diagnosis and treatment of eye diseases, solid tumours, inflammation caused by neovascularisation,
neovascular glaucoma, macular degeneration caused by neovascularisation,
diabetic retinopathy, macular degeneration, inflammatory bowel diseases, ulcers and
psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and
Crohn's disease.
Note: The sequence data for this patent did not form part of the printed

CC		Specification, but was obtained in electronic format directly from WIGO at ftp.wipo.int/pub/published_ptc_sequences.			
XX					
SQ	Sequence	9095 BP; 2977 A; 43 C; 1601 G; 4474 T; 0 other;			
	Query Match	2.8%; Score 46.2; DB 24; Length 9095;			
	Best Local Similarity	50.2%; P-adj. No. 0.085; Mismatches 113; Indels 0; Gaps 0;			
OY	MATCHES 114; Conservative	0; Mismatches 113; Indels 0; Gaps 0;			
DG	72	TGTTCACCAACAGAAATGACATTGCACCACCTCAGCATTTGCGAATAAGAATATTAGGAACACT 131 TATACACTTAACGAATTCATTTAACTATACCTTGCTTAATTAACAATCTACACACTAT 7428			
OY	132	TTTTTTTTTCTCTGAGTAGTATAGCAAAGAACGGGAAAAAAAACATTTCTTAGGGGG 191 TTTAAATCATTTATCCAATATATACATATTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 7368			
OY	192	ATTAACGGGCCAAGATACCAACAGCTTCAGAAAGAAAGATATATAGTAAGCAATATGAAAGA 251 AAA 7308			
OY	252	TAACTTCCTTTTCATCAAGAACGGCAAAAGTAATATATGATAGA 298 AAAAATATAA 7261			
DB	7307	AAAAATATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 7261			
	RESULT 13				
ID	AAL1189/C				
XX	AAL1189 standard; CDNA; 393 BP.				
XX	AAL1189;				
XX	07-DEC-2001 (first entry)				
DE	Human breast cancer expressed polynucleotide 3646.				
XX	Human; breast cancer; cell marker; cytosolic; ss.				
OS	Homo sapiens.				
XX	WO200151628-A2.				
PX	19-JUL-2001.				
PF	10-JAN-2001; 2001WCO-US00798.				
PR	14-JAN-2000; 2000US-0176077.				
PR	14-MAR-2000; 2000US-0189167.				
PR	24-MAR-2000; 2000US-0192099.				
PR	29-MAR-2000; 2000US-0193480.				
PR	15-MAY-2000; 2000US-0205230.				
PR	09-JUN-2000; 2000US-0211315.				
PA	25-JUL-2000; 2000US-0220534.				
(MIL-) MILLENNIUM PREDICTIVE MEDICINE INC.					
Lillie J, Xu Y, Wang Y, Steinmann K; WPI, 2001-451856/48. New peptide useful as a marker for the diagnosis of breast cancer - Claim 1; Page 671; 3695pp; English.					
The invention relates to human breast cancer expressed polynucleotides (AAID7544-AAID26789) and methods of assessing whether a patient is afflicted with breast cancer by examining the correlation between the expression of certain markers and the cancerous state of breast cells. The polynucleotides and encoded polypeptides are potential markers for detecting, diagnosing, monitoring, characterizing treating and potentially preventing breast cancer. The polynucleotides and encoded polypeptides are also useful for isolating compounds with cytostatic activity.					

[illegible]

Wed Feb 19 14:09:26 2003

us-09-605-042a-37.rst

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
February 17, 2003, 08:33:57 ; Search time 2083.49 Seconds
(without alignments)
12670.391 Million cell updates/sec

Run on:

US-09-605-042a-37

Title: 1630
Sequence: 1 actataggcagcagcgtggtc.....ccattctgttgaagatgg 1630

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
32308132

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 100%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*

1: em_estba:*
2: em_esthum:*
3: em_estinu:*
4: em_estnu:*
5: em_estov:*
6: em_estupl:*
7: em_estro:*
8: em_hic:*
9: qb_estl:*
10: qb_est2:*
11: qb_hic:*
12: qb_est13:*
13: qb_est4:*
14: qb_est5:*
15: em_estfun:*
16: em_estom:*
17: qb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pla:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_man:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the total score distribution, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	57.6	3.5	1101	17	CNS00396
2	56.2	3.4	813	17	CNS0119M
3	53.4	3.3	976	17	CNS018ND
4	52.6	3.2	413	14	BQ451939
5	52.4	3.2	903	17	CNS0011C
6	52.4	3.2	1045	17	CNS035E4

7	52.4	3.2	1065	17	CNS074RT
8	52.2	3.2	1101	17	CNS0029N
9	52.2	3.2	3040	10	BC025769
10	52	3.2	1300	10	BE420576
11	51.8	3.2	544	14	BQ398892
12	51.8	3.2	877	17	AZ211540
13	51.6	3.2	629	17	CNS04EYX
14	51.6	3.2	867	17	CNS074M9
15	51.6	3.2	1309	17	CNS016FX
16	51.4	3.2	1201	17	CNS01329
17	51	3.1	903	17	AZ54732V
18	50.8	3.1	840	17	CNS02WV
19	50.6	3.1	1058	17	CNS011HR
20	50.4	3.1	1201	17	CNS010TW
21	50.2	3.1	670	17	AV734180
22	50.2	3.1	920	17	CNS00372
23	50.2	3.1	920	17	AZ691914
24	50.2	3.1	994	17	BQ38127
25	50.2	3.1	1014	17	CNS04NOJ
26	50	3.1	1092	17	CNS06CAJ
27	50	3.1	1014	17	CNS020K7
28	50	3.1	679	9	AL106821
29	49.8	3.1	725	17	AL180166
30	49.8	3.1	996	17	CNS00PWH
31	49.6	3.0	239	14	BQ093422
32	49.4	3.0	841	12	BF267587
33	49.4	3.0	494	17	BH367087
34	49	3.0	536	14	BQ521883
35	49	3.0	557	14	BQ390979
36	49	3.0	887	17	CNS06OV4
37	49	3.0	887	17	BQ430188
38	48.8	3.0	331	17	CNS010C2
39	48.8	3.0	373	17	BH183499
40	48.8	3.0	494	14	BQ391155
41	48.8	3.0	1101	17	CNS00AYO
42	48.8	3.0	1101	17	CNS00K25
43	48.8	3.0	1101	17	CNS00LR2
44	48.8	3.0	1101	17	CNS04P4P
45	48.8	3.0	1184	17	CNS04P4P

ALIGNMENTS

RESULT 1
CNS00396
LOCUS
DEFINITION
CNS00396
Drosophila melanogaster genome survey sequence (fruit fly), genomic survey sequence.
1101 bp DNA linear GSS 03-JUN-1999
BACR08K10 of RPI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION
AL063921.1 GI:4941778
VERSION
GSS
KEYWORDS
Drosophila melanogaster.
SOURCE
Drosophila melanogaster
ORGANISM
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Phytolidae; Drosophilidae; Drosophila.

REFERENCE
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage
Submitted (02-JUN-1999) FRANCE (E-mail : seqrefgenoscope.cns.fr)
BP 191 91006 Evry cedex (fr)
AUTHORS
- Web : www.genoscope.cns.fr
TITLE
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP)
COMMENT
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org>
The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Goessawa and
Aaron Mammoler in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute by partial
NI. The library is named RPI-98 and was constructed by partial

score digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2: cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

location/Qualifiers

source

1. 1101

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone="BACR08K10"

/clone_11b="RPCI-98"

/note="end : TET3"

BASE COUNT 201 a 64 c 131 g 202 t 503 others

ORIGIN

Query Match 3.5%; Score 57.6; DB 17; Length 1101;

Best Local Similarity 18.3%; Pred. No. 0.0029;

Matches 125; Conservative 255; Mismatches 304; Indels 0; Gaps 0;

167 AAAAAAAAAACATCTTTATGGGGATTAACGGCAAGGATCAACAGCTGCAAAAG 226

418 AAAAAAAAAATATTAATAAATAAATAAATAAATAAATAAATAAATAAATAA 477

227 AATAAATAGTAAAGCAATGAAGAATACTCTTTTCATCAAGAACCGCAAGTAA 286

478 AAAAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 537

287 AATAATGATAGATCTTTCTCACTTTCCAAAGATGAAGATTAATGCGAGGCTG 346

538 ATAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 597

347 AGTACTGCTGGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 406

598 TGTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 657

407 GAAACATCGGTGAAGACGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 466

658 NAATAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 717

467 GAGCGGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 526

718 WTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 777

527 GCGATGCTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 586

778 WAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 837

587 TTTCTCAATGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 646

838 HWNTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 897

647 AGCGGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 706

898 NANTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 957

707 CACAGATGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 766

958 MNNTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 1017

767 ATTCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 826

1018 MTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1077

827 TACTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 850

1078 MAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1101

RESULT 2

CNS0119M 813 bp DNA linear GSS 26-JUL-1999

LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC

DEFINITION

BACN05K16 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.

AL099988.1 GI:5611599

VERSION

KEYWORDS

SOURCE

ORGANISM

Drosophila melanogaster.

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 813)

REFERENCE

AUTHORS

TITLE

JOURNAL

Direct Submission

Genoscope.

Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage;

BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

determination of this BAC-end sequence was carried out as part of a

collaboration with the European Drosophila Genome Project (EDGP) -

<http://www.edgp.ebi.ac.uk> - This Drosophila melanogaster BAC

library (Dros BAC) was made by Alain Billard at CEPH (Centre

d'Etude du Polymorphisme Humain) with funding provided by a MRC

project grant. The DNA was prepared from embryos by Alain Bucheton

and Genevieve Payan. It has been constructed in the vector

pBelobAC11.

location/Qualifiers

1. 813

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone="BACN05K16"

/clone_11b="DrosBAC"

/plasmid="pBelobAC11"

/note="end : T7"

BASE COUNT 434 a 83 c 81 g 93 t 122 others

Query Match 3.4%; Score 56.2; DB 17; Length 813;

Best Local Similarity 45.4%; Pred. No. 0.0062;

Matches 103; Conservative 22; Mismatches 102; Indels 0; Gaps 0;

105 GGATGAGAAAAGAAATATAGCACTTTATTTCTTCTGAAGTATAGCAAGAAAG 164

79 GCGNAGAAAGAAATCAAAAARATCATATAAATAAATAAAGAAATGAATAAAM 138

165 GGAATAAATAAATCTTTATGGGGATTAACGGCAAGGATCAACAGCTGCAAA 224

139 TGAATAAATAAATAAATTAATAAATAAATAAATAAATAAATAAATAAATAA 198

225 AGAATAAATAGTAAAGCAATGAAGAATACTCTTTTCATCAAGAACCGCAAG 284

199 AATTAAATATTAATAAARARARARARARARARARARARARARARARARAR 258

285 AATAATGATAGATGTTCTCACTTTTCCAAAGATGAAGATTAA 331

259 WACAAGCAAAAATAAATAATTTCAATTAATAAATAAATAAATAAATAAATAA 305

RESULT 3

CNS018HD 976 bp DNA linear GSS 26-JUL-1999

LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC

DEFINITION BACN13H03 of DrosBAC library from Drosophila melanogaster (fruit

fly), genomic survey sequence.

ACCESSION

VERSION

AL109339.1 GI:5629643

KEYWORDS

SOURCE

ORGANISM

Drosophila melanogaster.

Drosophila melanogaster.

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 976)

REFERENCE

Genoscope.

Submitted

Page 3

Washington University School of Medicine
Washington University Park Parkway, Box 8501, St. Louis, MO 63110
Tel: 314 286 1800
Fax 314 286 1810

REFERENCE
AUTHORS
TITLE
1 (basc)
Genoscope
Direct Submission

McDowell, J., Pearson, R., Strantrop, S., Thomas, P. J., Touchman, J. W., Tsurgoun, C., Vogt, J. L., Walker, M. A., Wetherby, K. D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E. D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 49 Row: 1 Column: 13
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA g1: 13376274
This clone has the following problem: frame shifted.

FEATURES

SOURCE

1. 3040
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5210832"
/tissue_type="Blood, adult leukocytes"
/clone_id="NIR_MGC_118"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
BASE COUNT 959 a 669 c 669 g 743 t
ORIGIN

Query Match 3.2%; Score 52.2; DB 11; Length 3040;
Best Local Similarity 53.1%; Pred. No. 0.06; Mismatches 0; Gaps 0;

Matches 111; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 35 GTAAAGACCCAGACTAGTTTACAGAGCCCTCATGTTCCACCAACAGAAATGACAT 94

DB 2830 GTAAAGAAAGAAATGATGTTTAAATAATACACCTTCATGAAATCAAGTGAT 2889

QY 95 TCACACCTAGATGAGAAAGAAATATAGCACTTTATTTCTTCGAAATTATG 154

DB 2890 TTTTCAGATTATGTCGCAAAATATATGTCATTCATTAATCTTGTAATAAAAAA 2949

QY 155 CAAGAAAGGGGAAATAAAACATTTCTATGGGATTAACGGCAAGATACAAAC 214

DB 2950 AA 3009

QY 215 AGTTCAGAAAGAAATTAATGTAACAAA 243

DB 3010 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3038

RESULT 10

BE420576

LOCUS

BE420576 1300 bp mRNA linear EST 24-JUL-2000

HMW000.D01 ITEC HMM Barley Leaf Library Hordeum vulgare cDNA clone

ACCESSION

VERSION

SOURCE

KEYWORDS

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Herrmann RG
Botanisches Institut der LMU
Menzinger Str. 67, D-80638 Munchen GERMANY
Fax: 49 30 171683
Email: herrmann@botanik.biologie.uni-muenchen.de
International Triticaceae EST Cooperative (ITEC)
<http://wheat.pw.usda.gov/genome>

FEATURES

SOURCE

Location/Qualifiers

1. 1300

/organism="Hordeum vulgare"

/cultivar="Barke"

/db_xref="taxon:4513"

/clone="HMW000.D01"

/tissue_type="Leaf"

/dev_stage="14 day old"

/note="Vector: pBluescriptSK(-); 850 bp average insert size."

BASE COUNT 566 a 207 c 214 g 272 t 41 others

ORIGIN

Query Match 3.2%; Score 52; DB 10; Length 1300;
Best Local Similarity 50.4%; Pred. No. 0.065;

Matches 127; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 80 AACCAAGAAATGACATTCACCACTTGAGATGAGAAAAAGAAATTTAGCACTTTATTT 139

DB 915 AAATAAAAAAATTTAAAAAATAAAAAAAAAAAAAAAAAAATAATTAATAATAA 974

QY 140 CTCTGAAATTAATGCAAGAAAGGGAATAAAACATTCCTTATGGGATTAACGG 199

DB 975 TAAAAAATAAAAAAAAAAAAAAAAAAATAATTAATAATAATAATAATAA 1034

QY 200 GCAAGGATCAAAACAGTTCCAGAAAGAAATTAATGTAAGCAATGAAATTAATTC 259

DB 1035 AAAAAAATAATTAATTAATAATAATAATAATAATAATAATAATAATAATAATA 1094

QY 260 TTTTTCATCAAGAACCCGCAAAAGTAATTAATGATTAAGTGTTCACCTTCCACAA 319

DB 1095 AAAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1154

QY 320 GATGAAGTTAA 331

DB 1155 TTTTAAAAAAA 1166

RESULT 11

BQ398892

LOCUS

BQ398892 544 bp mRNA linear EST 22-MAY-2002

NISC_m012b07.x1 Soares NXEG Xenopus laevis cDNA clone IMAGE:5278813

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgaaps-remail.nih.gov
cDNA Library Preparation: M. Bento Soares (University of Iowa)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC)
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1AM1704 row: D column: 14
Seq primer: -21M13 forward primer (ABI).
Location/Qualifiers

FEATURES

SOURCE

1. 544
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="IMAGE:5278813"

	/clone_1lb="G"	
	/note="Genoscope sequence ID : COBGI05CGI1ISPI-end :	
	pbc-Or1"	
BASE COUNT	352 a 51 c 60 g 108 t 58 others	
ORIGIN		
Query Match	3.2% ; Score 51.6 ; DB 17 ; Length 629 ;	
Best Local Similarity	44.7% ; Pred. No. 0.079 ;	
Matches 117 ;	Conservative 18 ; Mismatches 127 ; Indels 0 ; Gaps 0 ;	
OY 37	AAGACACCAGACTTATGGTTTGACAGACCTCATGTTCACCAACAGAATGACATTTC	96
Dd 39	AAAAACAATTTTATGGATAAATTTWAAAAGCCAAATCTCACAAACAAAACAAAGCAATTTT	98
OY 97	ACCACCTAGATTGAGAAAAAGAAATATTAGGAACCTTTATTTCTCTGTGAAGTTATAGCA	156
Dd 99	WTTCGTGMAAAATATTAGGAAAAARAAAAAATAAAAAATTTGTAATWAAAAACAATATWAA	158
OY 157	AAGAAAGGGAAAAAANAACATCTTATGGGGGTAAACGGCAGAAAGATACAAACAG	216
Dd 159	GAAAAAATTT	218
OY 217	TTCAGAAAAAGAAATTAATAGTAGAACAAATGAAAAAGATTAACCTCTTTTTCATCAAAGAAC	276
Dd 219	AACTTAMTTWTTAAAAA	278
OY 277	GCAAAAGTAATTAATGATTAAGA 298	
Dd 279	AAAAAAAAAAAAAAAAAAAAAARGAARA 300	
RESULT 14		
CNS074W9		
LOCUS		
DEFINITION	CNS074W9 867 bp DNA linear GSS 07-JUL-2001	
ACCESSION	clone BA0AB031B08 of library BA0AB from strain CLIB 210 of	
VERSION	Kluyveromyces lactis, genomic survey sequence.	
KEYWORDS	Ala29343	
SOURCE	Ala29343.1 GI:12212537	
ORGANISM	GSS.	
REFERENCE	Kluyveromyces lactis.	
AUTHORS	Kluyveromyces lactis	
	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;	
	Saccharomycetales; Saccharomycetaceae; Kluyveromyces.	
	1 (bases 1 to 867)	
	Soulet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,	
	Bolotin-Fukuhara,M., Bon,E., Broctier,P., Casaregola,S.,	
	de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Florente,B.,	
	Malpertuy,A., Neuvéglise,C., Ozier-Kalogiropoulos,O., Potier,S.,	
	Saurin,W., Tekra,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,	
	Wincker,P. and Weissenbach,J.	
	Genomic exploration of the hemiascomycetous yeasts: I. A set of	
	yeast species for molecular evolution studies	
	FEBS Lett. 487 (1), 3-12 (2000)	
TITLE	11152876	
JOURNAL	2 (bases 1 to 867)	
MEDLINE	Bolotin-Fukuhara,M., Toffano-Nioche,C., Artiguenave,F.,	
PUBMED	Duchateau-Nguyen,G., Lemaire,M., Marmelise,R., Montrocher,R.,	
REFERENCES	Robert,C., Termier,M., Wincker,P. and Wesolowski-Louvel,M.	
AUTHORS	Genomic exploration of the hemiascomycetous yeasts: II.	
	Kluyveromyces lactis	
	FEBS Lett. 487 (1), 66-70 (2000)	
TITLE	11152886	
JOURNAL	20584721	
MEDLINE	FESB Lett. 487 (1), 66-70 (2000)	
PUBMED	11152886	
REFERENCES	3 (bases 1 to 867)	
AUTHORS	Genoscope.	
TITLE	Direct Submission	
JOURNAL	Submitted (08-SEP-2000) Genoscope - Centre National de Sequençage,	
	2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :	
	secref@genoscope.cns.fr Web : www.genoscope.cns.fr)	
	This GSS is part of a random genomic sequencing program of thirteen	
	yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces	
	exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,	

Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces fragilis var. fragilis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

location/Qualifiers

1..867

/organism="Kluyveromyces fragilis"

/strain="CLIB 210"

/varietal="fragilis"

/db_xref="taxon:28985"

/clone="BA0AB031B08"

/clone_1id="BA0AB"

BASE COUNT 292 a 81 c 128 g 216 t 150 others

ORIGIN

Query Match 3.28; Score 51.6; DB 17; Length 867;

Best Local Similarity 46.8%; Pred. No. 0.08;

Matches 96; Conservative 16; Mismatches 93; Indels 0; Gaps 0;

OY 103 TAGGATTTAGCAAAAGATATTAGCAACTTATTTCTTCGAAGTTATAGCAACAAA 162
| : : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : :
Db 640 TTTTMMWWWWWATWTAATTTTWTANAMAWATTTTTTTTAAATTTNNAAAAA 699
| : : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : :
OY 163 GGGGAAAAAAAAAACATCTTTATGGGGATPAAGCGCAAAGATACAACAGTTTCCA 222
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 700 AAAAAAAAAAAAAAAAAANTTTATTTWMTATTTAAAAATATWATATATTTAAAAAGAA 759
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 223 AAAAGATAAATAGTAAACCAAGAAGATTAATCTCCTTTTCATCAAGAACCCGCAAAA 282
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 760 AAAAAAAAAATTAAMWAAMAAATATTTATTAATAAAAMWTTATTTTWTATAAAATAATAAAA 819
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 283 GTAAATATATGATTAAGATGTTCCTCA 307
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 820 AAATAAATATATATTTATTTATATAA 844
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 15
BE420736/c

LOCUS
DEFINITION HMM002.A05 ITFC HM Barley Leaf Library Hordeum vulgare cDNA clone
HMM002.A05, mRNA sequence.

ACCSSION
VERSION BE420736
KEYWORDS BE420736.1 GI:9418579
EST.

SOURCE
ORGANISM Hordeum vulgare.
SOURCE Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
Triticeae; Hordeum.
1 (bases 1 to 1309)

REFERENCE
AUTHORS Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier,
S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,
Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,
Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y.,
Pecchiaroni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M.,
Sorrells,M., Warburton,M. and Wenzel,G.
International Triticeae EST Cooperative (ITFC) : Production of
Expressed Sequence Tags for Species of the Triticeae
Unpublished (2000)
Contact: Herrmann RG
Botanisches Institut der LMU
Menzinger Str. 67, D-80638 München GERMANY
Fax: 49 30 171683
Email: herrmann@botanik.biologie.uni-muenchen.de
International Triticeae EST Cooperative (ITFC)
http://wheat.pw.usda.gov/genome.
Location/Qualifiers

1..1309

/organism="Hordeum vulgare"

/cultivar="Barke"

Wed Feb 19 14:09:26 2003

us-09-605-042a-37.rst

```
/db_xref="taxon:4513"
/clone="HM002.A05"
/clone_id="JREC HMW Barley Leaf Library"
/tissue_type="leaf"
/dev_stage="14 day old"
/Note="Vector: pBluescriptSK(-); 850 bp average insert
size: 172 c 159 g 615 t 156 others"

BASE COUNT      207 a 172 c 159 g 615 t 156 others
ORIGIN
Query Match      3.2%: Score 51.6; DB 10; Length 1309;
Best Local Similarity 44.1%; Pred. No. 0.082; Mismatches 171; Indels 0; Gaps 0;
Matches 135; Conservative
OY 80 ACCAGAAATGACATTCACCCAGTGTGAGAAAAAGAAATTTAGACCTTTTATTTT 139
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 117 AATGAAAAAATGAAAAAATGAAAAAATGAAAAAATGAAAAAATGAAAAAATG 1058
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 140 CTCTGAGTATATAGCAAGAAAGGGAACATCTATGCGGATTAACGG 259
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1117 NTTTTNTTTTAAAAAATGAAAAAATGAAAAAATGAAAAAATGAAAAAATG 998
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 200 GCAAGATGCAACAGTTCAGAAAAAGATTAATGATGACCAATGAAAGATACTTC 319
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1057 ATTTAAATTAATAATTAATAATTAATAATTAATAATTAATAATTAATAAT 938
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 260 TTTTCATCAAGACCGCAAAAGTAATATGATAGATGTTCTCCTTCCACAAA 379
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 997 AAAAAAATTAATGCGCCAGGCTGAGTACTGCTGGGATTTGAACTAATCTTAT 878
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 320 GATGAAGTATATGCGCCAGGCTGAGTACTGCTGGGATTTGAACTAATCTTAT 878
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 937 AAAAAAATTAATAATTAATAATTAATAATTAATAATTAATAATTAATAAT 878
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 380 AGATCT 385
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 877 TTAAT 872
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Search completed: February 17, 2003, 16:07:13
Job time : 2104.74 secs
```


us-09-605-042a-37.rge

Wed Feb 19 14:09:20 2003

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
February 17, 2003, 08:20:43 ; Search time 4070.6 seconds
11653.717 Million cell updates/sec

US-09-605-042a-37

Perfect score: 1630
Sequence: 1 actataggcgcgcgtgctc.....ccattctgttgaagatcg 1630

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues 4109280

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: gb_da:*
- 2: gb_hcg:*
- 3: gb_in:*
- 4: gb_ov:*
- 5: gb_ov:*
- 6: gb_ov:*
- 7: gb_pl:*
- 8: gb_pl:*
- 9: gb_pl:*
- 10: gb_pl:*
- 11: gb_pl:*
- 12: gb_pl:*
- 13: gb_pl:*
- 14: gb_pl:*
- 15: gb_pl:*
- 16: gb_pl:*
- 17: gb_pl:*
- 18: gb_pl:*
- 19: gb_pl:*
- 20: gb_pl:*
- 21: gb_pl:*
- 22: gb_pl:*
- 23: gb_pl:*
- 24: gb_pl:*
- 25: gb_pl:*
- 26: gb_pl:*
- 27: gb_pl:*
- 28: gb_pl:*
- 29: gb_pl:*
- 30: gb_pl:*
- 31: gb_pl:*
- 32: gb_pl:*
- 33: gb_pl:*
- 34: gb_pl:*
- 35: gb_pl:*
- 36: gb_pl:*
- 37: gb_pl:*
- 38: gb_pl:*
- 39: gb_pl:*
- 40: gb_pl:*
- 41: gb_pl:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	557.6	34.2	626	4 S75961	S75961 TTP-Tamm-Ho
2	373.4	22.9	168887	9 AC106796	AC106796 Homo sapi
3	372.4	22.8	6727	9 AY061638	AY061638 Homo sapi
4	269.2	16.5	162548	2 AC015831	AC015831 Homo sapi
5	269.2	16.3	620	9 S75968	S75968 TTP-Tamm-Ho
6	202.4	12.4	162548	2 AC015831	AC015831 Homo sapi
7	147.6	9.0	165503	2 AC115158	AC115158 Mus muscu
8	143.6	8.8	219386	2 AC125147	AC125147 Mus muscu
9	141.4	8.7	625	10 S75965	S75965 TTP-Tamm-Ho
10	54.6	3.3	82392	2 AC117176	AC117176 Dictyoste
11	53.6	3.3	82392	2 AC118290	AC118290 Rattus no
12	52.4	3.2	259782	2 AC115218	AC115218 Rattus no
13	52.2	3.2	131665	2 AC115680	AC115680 Dictyoste
14	51.4	3.2	59004	2 AC1084799	AC1084799 Mus muscu
15	51.2	3.2	303091	2 AC117080	AC117080 Dictyoste
16	51.2	3.1	106434	3 AC117074	AC117074 Dictyoste
17	50.8	3.1	61052	2 AC112513	AC112513 Homo sapi
18	50.8	3.1	170536	9 AC093914	AC093914 Homo sapi
19	50.8	3.1	184425	9 AC025593	AC025593 Homo sapi
20	50.6	3.1	2418	3 DDBASG6	DDBASG6 Mus muscu
21	49.8	3.1	219626	10 AC098719	AC098719 Mus muscu
22	49.2	3.0	45210	2 AC087043	AC087043 Homo sapi
23	48.2	3.0	16685	2 AC107555	AC107555 Rattus no
24	48.2	3.0	206456	2 AC112402	AC112402 Homo sapi
25	48.6	3.0	179456	2 AC023940	AC023940 Plasmid
26	48.6	3.0	205429	9 AC005506	AC005506 Human chr
27	48.6	3.0	225603	2 AC001866	AC001866 Homo sapi
28	48.6	3.0	239685	2 AC087242	AC087242 Homo sapi
29	48.6	3.0	333530	2 AC066310	AC066310 Rattus no
30	48.6	3.0	89004	3 CER6942	CER6942 Rattus no
31	48.4	3.0	103093	2 AC068636	AC068636 Homo sapi
32	48.4	3.0	136519	2 AC068636	AC068636 Homo sapi
33	48.4	3.0	154809	2 AC061890	AC061890 Homo sapi
34	48.4	3.0	168011	9 AC061890	AC061890 Homo sapi
35	48.4	3.0	168011	9 AC061890	AC061890 Homo sapi
36	48.4	3.0	172032	2 AC021322	AC021322 Homo sapi
37	48.4	3.0	172032	2 AC021322	AC021322 Homo sapi
38	48.4	3.0	195430	2 AC048333	AC048333 Homo sapi
39	48.4	3.0	292767	2 AC026347	AC026347 Rattus no
40	48.2	3.0	110000	2 AC013555-0	AC013555-0 Rattus no
41	48.2	3.0	300605	2 AC098460	AC098460 Rattus no
42	48.2	2.9	152053	2 AC112413	AC112413 Rattus no
43	48.2	2.9	152053	2 AC107092	AC107092 Rattus no
44	47.8	2.9	151371	2 AC091889	AC091889 Homo sapi
45	47.8	2.9	151371	2 AC091889	AC091889 Homo sapi

ALIGNMENTS

RESULT 1
LOCUS S75961
DEFINITION TTP-Tamm-Horsfall protein [promoter] [cattle, Genomic, 626 nt].
ACCESSION S75961
VERSION S75961.1 GI:912818
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE

Pred No is the number of results predicted by chance to have a

FEATURES		Location/Qualifiers	
gene	1.. 6727	/organism="Homo sapiens"	
promoter	/db_xref="taxon:9606"		
mRNA	/chromosome="16"		
5'UTR	1.. >6727		
CDS	/gene="UMOD"		
	1.. 3706		
	/gene="UMOD"		
	join(3707.. 3731,5646.. >5772)		
	/gene="UMOD"		
	/product="uromodulin"		
	3707.. 5684		
	/gene="UMOD"		
	5685.. >5772		
	/gene="UMOD"		
	/note="transmembrane glycoprotein"		
	/codon_start=1		
	/product="uromodulin"		
	/protein_id="AA127853.1"		
	/db_xref="gi:20384858"		
	/translation="MGPSLTLWMLKRYVASFITTAADTSEA"		
	/translation="MGPSLTLWMLKRYVASFITTAADTSEA"		
	1420 c 1552 g 1822 t		

BASE COUNT	1933 a	1420 c	1552 g	1822 t
ORIGIN	22.8%;	Score 372.4;	DB 9;	Length 6727;
Query Match	Similarity 67.7%;	Pred. No. 36-81;	DB 9;	Length 6727;
Best Local	Matches 703;	Conservative	Mismatches 286;	Indels 49;
				Gaps 11;

ORIGIN	1933 a	1420 c	1552 g	1822 t
590	CTCAATATGCAATGGCATTACGTAGTGAATAGGCTCAGAGACCTTAAGCTGACCAACG	649		
2750	CTGACTTTTGAATGTCATGTATGTTGAACTGGCTGAAACATGTTTACTGAGCC	2809		
650	CCTGCACTTCCTGCCCCA--TGACTGGCATCACCACCAAGCATCGGCAGTACTCC	707		
2810	TCTGTCACTTTTACCTCCAGAGTTCACACACAGCAAGCAGAAATTTTCACT-ACCTTC	2868		
708	ACAGATGTCACATGCTATGATAGCTTAAGCTTAAGCTTAAGCTTAAGCTTAAGCTTAAGCT	2928		
2869	ATGAAATATATCACTATCACTATGATGCTTAAGCTTAAGCTTAAGCTTAAGCTTAAGCT	2988		
768	TTCATAGGACAGGGTATTTCTTTTACAGATGATGAATTAATTAATTAATTAATTAATTAAT	872		
2929	CTGTGGAGATGGGTATTCATTTTATGACAAAGAAATTAATTAATTAATTAATTAATTAAT	3048		
818	----AAGGAAATTAAGTGCACACAAAAAACCTGACGACATTCAGATGACTGCTC	929		
2989	GATTAAATATGAGCCACTGACACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	3108		
873	CAAACTCAGTCTGATCACTACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	989		
3049	CAAACTCAGTCTGATCACTACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	3168		
930	GATTCAATTTCTTGATTCACAGAGCATCGTGTCCATGATGTCGAAATTAATTAATTAATTAAT	1049		
3109	GACCTGATTTCTTTGTTGACAGAGGGGTGACGTCACATGATGTCGAAATTAATTAATTAAT	3221		
990	CTTGACCTCAGACATTCCTCAGCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	1106		
3169	TTCCTACTTTCAGCCATTCACACT	3281		
1050	GCTGAATCTGAGAGATGGCATTTGCTTTGGAATTAATTAATTAATTAATTAATTAATTAAT	1166		
3222	GCTGAATCTGAGAGATGGCATTTGCTTTGGAATTAATTAATTAATTAATTAATTAATTAAT	3341		
1107	TGTTGGAGACATGAGACACCCCTTTTCTCCGAGAGACCTGAGCTCAGATTCCTGCTC	3381		
3282	CATCTGGGGCAGTCAATATCCCTTTTGCTTATGACAGCCCTGGCTTTCAGTACTGCTC	3441		

Wed Feb 19 14:09:20 2003

us-09-605-042a-37.1rg

```

1167 TGCTGGCT-----CCACTTGTGCTTTCAATGACCAAGAA-ATCCGAGCCCTT 1217
OY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3342 TGCCGCTCTCTGATCCACCTGCTGACGTGACAGAGAGCCGACCTT 3401
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1218 GGAATGTTACTCACTTAATTTCTACTAAGAACCTTTGTCACAAAGATATAAA 1277
OY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3402 GGCACGCTTCCGATTAATTTCTACTAGATCTCTGCTGTAGAGGCGCAAA 3461
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1278 CAGAGCCCTTGTAGCTGTGGGACAGCTGTGACCCC--ATGTCATCTTTGGGCTCT 1335
OY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3462 CAGTACCTTGTATTCGGGACAGGTGTACCCCGCCCATGTCAATCTTTGGGCTCT 3521
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1336 TACTATTAGGAAAGACACACACACCTCCACCTAGAGAGAGAAACCTGTGTC 3581
OY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3522 TAGCTTTAGGAAAGACACACACCTCCACCTAGAGAGAGAAACCTGTGTC 3641
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1396 AAGAGGAAATATTCACCCCTTAATAA--ATTAGAAACAGACAGAGATCA 1453
OY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3582 AGGAGGAAATATTCACCCCTTAATAA--ATTAGAAACAGACAGAGATCA 1513
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1454 TTGAGAGAGATTCAGCTGGGACAGATGTATATATATATGAAAGTCACTAC 1513
OY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3642 TTGGAGGAGATTCAGCTGGGACAGATGTATATATATGAAAGTCACTAC 1573
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1514 TTGTAAGAGATTAATCTTACCTTTGCTTTCAGTAAAGCTATCTGAGCTCT 3756
OY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3697 TCATGACGACCTACTACTCTTCTGCTTCAGTAAAGCTATCTGAGCTCT 3756
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1574 CTCCTAGACCACTTCC 1591
OY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3757 CCTAGCCAGCTTCTCC 3774
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 4
AC015831 162548 bp DNA Linear HTG 26-MAR-2000
LOCUS Homo sapiens clone RP11-688624, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC015831.2 GI:7229398
ACCESSION AC015831.2
VERSION HTGS-PHASE0.
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE 1. Birren B., Linton L., Nusbaum C. and Lander E.
1 (bases 1 to 162548)
Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
REFERENCE 2. Birren B., Linton L., Nusbaum C., Lander E., Allen N., Anderson M.,
1 unpublished
2 (bases 1 to 162548)
Baldwin, J., Castle, A., Colangelo, M., Collins, S., Collymore, A., Doyle, M.,
Baldwin, J., Castle, A., Colangelo, M., Collins, S., Collymore, A., Doyle, M.,
Brown, A., Deaellano, K., Dwyer, K., Dominko, M., Dwyer, K., Gage, D., Horton, L.,
Cook, P., FitzHugh, W., Forrest, C., Funke, H., Haefford, A., Klein, J.,
Ferreira, P., Gardina, S., Grant, C., Kann, L., Marquis, N.,
Galagan, J., Johnson, R., Jones, C., Kann, L., Marquis, N.,
Howard, J., Liu, C., Locke, A., McKernan, K., McDonald, P., Melidri, J.,
Lehoczky, J., McMur, A., McKernan, K., McDonald, P., Melidri, J.,
McEwan, P., Maylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
Morrow, J., Pollara, V., Riley, R., Santoro, R., Santos, R., Severy, J.,
Peterson, K., Stojanovic, N., Subramanian, A., Talamas, J.,
Stange, Rhomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Teafaye, S., Tirelli, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 26, 2000 this sequence version replaced gi:16446837.
All repeats were identified using RepeatMasker.html
Smit, A. F. A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center/ MIT Center for Genome Research
Center Whitehead Institute/ MIT Center for Genome Research

```

```

Center code: WTB8
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: 14338
Center clone name: 688_G_24
-----
NOTE: This record contains 173 individual
sequencing reads that have not been assembled into
contigs. Runs of N are used to separate the reads
and the order in which they appear is useful for
arbitrary. Low-pass sequence sampling and allows
identifying relationships among clones that this clone
overlap. It should not be assumed that the event that
will be sequenced to completion. In the event that
the record is updated, the accession number will
be preserved.
858: contig of 858 bp in length
859: gap of 100 bp
859: 958: gap of 876 bp in length
959: 1834: contig of 100 bp
1835: 1934: gap of 100 bp
1935: 2832: contig of 898 bp in length
2833: 2932: gap of 100 bp
2933: 3789: contig of 857 bp in length
3790: 3889: gap of 100 bp
3890: 4780: contig of 891 bp in length
4781: 4880: gap of 100 bp
4881: 5740: contig of 860 bp in length
5741: 5840: gap of 100 bp
5841: 6742: contig of 902 bp in length
6743: 6842: gap of 100 bp
6843: 7684: gap of 100 bp
7685: 7784: gap of 100 bp
7785: 8603: gap of 100 bp
8610: 8709: gap of 100 bp
8710: 9573: gap of 100 bp
9574: 9673: gap of 864 bp in length
9674: 10537: contig of 864 bp in length
10538: 10637: gap of 100 bp
10638: 11521: gap of 884 bp in length
11522: 11621: gap of 100 bp
11622: 12466: contig of 845 bp in length
12467: 12566: gap of 100 bp
12567: 13464: contig of 898 bp in length
13465: 13564: gap of 100 bp
13565: 14419: contig of 855 bp in length
14420: 14519: gap of 100 bp
14520: 15368: contig of 849 bp in length
15369: 15468: gap of 100 bp
15469: 16325: contig of 857 bp in length
16326: 16425: gap of 100 bp
16426: 17261: contig of 836 bp in length
17262: 17361: gap of 100 bp
17362: 18185: contig of 824 bp in length
18186: 18285: gap of 100 bp
18286: 19098: contig of 813 bp in length
19099: 19198: gap of 100 bp
19199: 20057: contig of 859 bp in length
20058: 20157: gap of 100 bp
20158: 21021: gap of 864 bp in length
21022: 21121: gap of 100 bp
21122: 21995: contig of 874 bp in length
21996: 22095: gap of 100 bp
22096: 22965: contig of 870 bp in length
22966: 23065: gap of 100 bp
23066: 23957: contig of 892 bp in length
23958: 24057: gap of 100 bp
24058: 24926: contig of 869 bp in length
24927: 25026: gap of 100 bp
25027: 25881: contig of 855 bp in length
25882: 25981: gap of 100 bp
25982: 26827: contig of 846 bp in length
25982

```

26828	26927:	gap of	100 bp	in length
26928	27791:	contlg of 864 bp	100 bp	in length
27792	27891:	gap of 100 bp	100 bp	in length
27892	28742:	contlg of 851 bp	100 bp	in length
28743	28842:	gap of 100 bp	100 bp	in length
28843	29689:	contlg of 847 bp	100 bp	in length
29690	29789:	gap of 100 bp	100 bp	in length
29790	30642:	contlg of 853 bp	100 bp	in length
30643	30744:	gap of 100 bp	100 bp	in length
30743	31640:	contlg of 898 bp	100 bp	in length
31641	31740:	gap of 100 bp	100 bp	in length
31741	32602:	contlg of 862 bp	100 bp	in length
32603	32702:	gap of 100 bp	100 bp	in length
32703	33554:	contlg of 852 bp	100 bp	in length
33555	33654:	gap of 100 bp	100 bp	in length
33655	34521:	contlg of 867 bp	100 bp	in length
34522	34621:	gap of 100 bp	100 bp	in length
34622	35483:	contlg of 862 bp	100 bp	in length
34623	35583:	gap of 100 bp	100 bp	in length
34624	35883:	contlg of 872 bp	100 bp	in length
35884	36455:	contlg of 100 bp	100 bp	in length
35885	36555:	gap of 100 bp	100 bp	in length
36456	37144:	contlg of 859 bp	100 bp	in length
36457	37144:	contlg of 100 bp	100 bp	in length
37145	37514:	gap of 100 bp	100 bp	in length
37515	38355:	contlg of 841 bp	100 bp	in length
38356	38455:	gap of 100 bp	100 bp	in length
38456	39286:	contlg of 843 bp	100 bp	in length
39289	39386:	gap of 100 bp	100 bp	in length
39399	40233:	contlg of 836 bp	100 bp	in length
40235	40333:	gap of 100 bp	100 bp	in length
40335	41179:	contlg of 845 bp	100 bp	in length
41180	41273:	gap of 100 bp	100 bp	in length
41280	42148:	contlg of 869 bp	100 bp	in length
42149	42284:	gap of 100 bp	100 bp	in length
42289	43053:	contlg of 833 bp	100 bp	in length
43084	43183:	gap of 100 bp	100 bp	in length
43184	44082:	contlg of 899 bp	100 bp	in length
44083	44182:	gap of 100 bp	100 bp	in length
44183	45028:	contlg of 846 bp	100 bp	in length
45029	45128:	gap of 100 bp	100 bp	in length
45129	45593:	contlg of 865 bp	100 bp	in length
45594	46093:	gap of 100 bp	100 bp	in length
46094	46970:	contlg of 877 bp	100 bp	in length
46971	47070:	gap of 100 bp	100 bp	in length
47072	47976:	contlg of 906 bp	100 bp	in length
47977	48076:	gap of 100 bp	100 bp	in length
48077	48933:	contlg of 857 bp	100 bp	in length
48934	49033:	gap of 100 bp	100 bp	in length
49034	49927:	contlg of 894 bp	100 bp	in length
49928	50027:	gap of 100 bp	100 bp	in length
50028	50811:	contlg of 844 bp	100 bp	in length
50812	50971:	gap of 100 bp	100 bp	in length
50972	51816:	contlg of 845 bp	100 bp	in length
51817	51916:	gap of 100 bp	100 bp	in length
51917	52755:	contlg of 839 bp	100 bp	in length
52756	52855:	gap of 100 bp	100 bp	in length
52856	53707:	contlg of 852 bp	100 bp	in length
53708	53807:	gap of 100 bp	100 bp	in length
53808	54687:	contlg of 880 bp	100 bp	in length
54688	54787:	gap of 100 bp	100 bp	in length
54788	55651:	contlg of 864 bp	100 bp	in length
55652	55751:	gap of 100 bp	100 bp	in length
55752	56608:	contlg of 857 bp	100 bp	in length
56609	56708:	gap of 100 bp	100 bp	in length
56709	57576:	contlg of 868 bp	100 bp	in length
57577	57676:	gap of 100 bp	100 bp	in length
57677	58531:	contlg of 855 bp	100 bp	in length
58532	58631:	gap of 100 bp	100 bp	in length
58633	59474:	contlg of 843 bp	100 bp	in length
59475	59574:	gap of 100 bp	100 bp	in length
59575	60408:	contlg of 834 bp	100 bp	in length
60409	60508:	gap of 100 bp	100 bp	in length
60509	61396:	contlg of 888 bp	100 bp	in length
61397	61496:	gap of 100 bp	100 bp	in length

Query Match Similarity 71.0%; Prev. Mismatches 164; nuc

Best local similarity: conservative

Matches 461: conservative

XY 871 TCCAAAGTCAGCTCTGCTCATCTACATCAAAATTAATCTACTTCTCAAGACCTTCC--TTCATGG 929

Db 108315 TCAAAGTCAGCTATTTAAAGTCAAATTAATTTTGGCTTCCAAAGGAGACTCCCACTCA 108256

XY 930 GAATTCAAATTTCTGATTCACAGAGCATCTGTCTCCAAATGATCTGAAATTCTCGCTGT 989

Db 108255 GAGCTGATCTTTTGTGTATACAGAGTGGGTGACAGTCTGATGTCTGTACACTTCTGG 1049

XY 990 CTCCTACCTTCAGCCATTCACCTCAGCTCTTCTCTGTATCATCATTTGGACCCCAAGGAGCTG 108143

Db 108195 TTCTGACTTTCAGCCATTCCTAGCTCTCTCTCTGCTTGTGTGTGATCTCTA-----AG 108143

XY 1050 GCTGATCTGTGTGAGATGGCATTTGCTTGGAAATTAATGAGCC---ACAATGACATATCC 1106

Db 108142 GCTGATCTGATGAGATGGGTGTTTCACAAAGGTGGCCCTCTCCAAACAGAGTGCACCTTC 108083

XY 1107 TGTGTGGAGCATGAGACACCCCTTTCTCTCTGTGAGACACCTCTGGCTTCAGATTTGCGCTC 1166

Db 108082 CATCTGGGCACTGATATATCCCTTTGCTCTTATCATGACCAAGAA--AATCCAGGCCCTT 1217

XY 1167 TCGTTGAGCT-----CCACTTTTGCTCTTTCATATGACCAAGAA--AATCCAGGCCCTT 107963

Db 108022 TGCCCTGGCTCTTGATCCCAACCCCTGCTGTGTGATGACCAAGAAACCCAGACCTT 1277

XY 1218 GGAATGTTTACCTACGATTAATTTCTACTAAAGAACCTCTGTTGCCCAAAAGGTATMAA 107980

Db 107962 GGCACCTGCTTCCAGTTAATTTCTACATGGAATCTCTTGCTGTATGAAGGTGGGAAA 107983

XY 1278 CAGAGCCCTTGTAGCTGTGGGACAGCTGTGACCCCC--ATGTCAATCATTTGGGGCTTC 1335

Db 107902 CAGTGAACCTTGATTTCCGGGACAGAGTGTGACCCCCCAATGTCAATCATTTGGGGCTTC 107843

XY 1336 TACCTATTAGGAAAGAAACAACAACACACTCAGCTGTGGACAAAGGCAAACTATTATGCG 107783

Db 107842 TAGCTATTAGGAAAGAAACAACAACACACTCAGCTGTGGACAAAGGCAAACTATTATG 107723

XY 1396 AAAAGGAAATATTTCCACCCCTTAAATAA--ATTAGGAACAGAGACAGAGATCA 1453

Db 107782 AGAGAGAAAAATATTTCCACCCCAAGAAACAAATATCCAAAAAGCAATGAGACTAA 107723

XY 1454 TTGGAGAGAGATTTGCGAGTGGGGCAAGATGTATATTAATAGTATGA 107674

Db 107722 TTGGAGAGAGATTTGCGAGTGGGGCAAGATGTATATTAATAGTATGA 107674

RESULT 5 620 bp DNA linear PRI 27-JUL-1995

LOCUS S75968 TBP-Tamm-Horsfall protein [promoter] [human, genomic, 620 nt]

DEFINITION S75968

ACCESSION S75968.1 GI:912820

VERSION


```
*      6843      7684: contig of 842 bp in length
*      7685      7784: gap of 100 bp
*      7785      8609: contig of 825 bp in length
*      8610      8709: gap of 100 bp
*      8710      9573: contig of 864 bp in length
*      9574      9673: gap of 100 bp
*      9674      10537: contig of 864 bp in length
*      10538      10637: gap of 100 bp
*      10638      11521: contig of 884 bp in length
*      11522      11621: gap of 100 bp
*      11622      12466: contig of 845 bp in length
*      12467      12566: gap of 100 bp
*      12567      13464: contig of 898 bp in length
*      13465      13564: gap of 100 bp
*      13565      14419: contig of 855 bp in length
*      14420      14519: gap of 100 bp
*      14520      15368: contig of 849 bp in length
*      15369      15468: gap of 100 bp
*      15469      16325: contig of 857 bp in length
*      16326      16425: gap of 100 bp
*      16426      17261: contig of 836 bp in length
*      17262      17361: gap of 100 bp
*      17362      18185: contig of 824 bp in length
*      18186      18285: gap of 100 bp
*      18286      19098: contig of 813 bp in length
*      19099      19198: gap of 100 bp
*      19199      20057: contig of 859 bp in length
*      20058      20157: gap of 100 bp
*      20158      21021: contig of 864 bp in length
*      21022      21121: gap of 100 bp
*      21122      21995: contig of 874 bp in length
*      21996      22095: gap of 100 bp
*      22096      22965: contig of 870 bp in length
*      22966      23065: gap of 100 bp
*      23066      23957: contig of 892 bp in length
*      23958      24057: gap of 100 bp
*      24058      24926: contig of 869 bp in length
*      24927      25026: gap of 100 bp
*      25027      25881: contig of 855 bp in length
*      25882      25981: gap of 100 bp
*      25982      26827: contig of 846 bp in length
*      26828      26927: gap of 100 bp
*      26928      27791: contig of 864 bp in length
*      27792      27891: gap of 100 bp
*      27892      28742: contig of 851 bp in length
*      28743      28842: gap of 100 bp
*      28843      29689: contig of 847 bp in length
*      29690      29789: gap of 100 bp
*      29790      30642: contig of 853 bp in length
*      30643      30742: gap of 100 bp
*      30743      31640: contig of 898 bp in length
*      31641      31740: gap of 100 bp
*      31741      32602: contig of 862 bp in length
*      32603      32702: gap of 100 bp
*      32703      33554: contig of 852 bp in length
*      33555      33654: gap of 100 bp
*      33655      34521: contig of 867 bp in length
*      34522      34621: gap of 100 bp
*      34622      35483: contig of 862 bp in length
*      35484      35583: gap of 100 bp
*      35584      36455: contig of 872 bp in length
*      36456      36555: gap of 100 bp
*      36556      37414: contig of 859 bp in length
*      37415      37514: gap of 100 bp
*      37515      38355: contig of 841 bp in length
*      38356      38455: gap of 100 bp
*      38456      39298: contig of 843 bp in length
*      39299      39398: gap of 100 bp
*      39399      40234: contig of 836 bp in length
*      40235      40334: gap of 100 bp
*      40335      41179: contig of 845 bp in length
*      41180      41279: gap of 100 bp
*      41280      42148: contig of 869 bp in length
```

```
*      42149      42248: gap of 100 bp
*      42249      43083: contig of 835 bp in length
*      43084      43183: gap of 100 bp
*      43184      44082: contig of 899 bp in length
*      44083      44182: gap of 100 bp
*      44183      45028: contig of 846 bp in length
*      45029      45128: gap of 100 bp
*      45129      45993: contig of 865 bp in length
*      45994      46093: gap of 100 bp
*      46094      46970: contig of 877 bp in length
*      46971      47070: gap of 100 bp
*      47071      47976: contig of 906 bp in length
*      47977      48076: gap of 100 bp
*      48077      48933: contig of 857 bp in length
*      48934      49033: gap of 100 bp
*      49034      49927: contig of 894 bp in length
*      49928      50027: gap of 100 bp
*      50028      50871: contig of 844 bp in length
*      50872      50971: gap of 100 bp
*      50972      51816: contig of 845 bp in length
*      51817      51916: gap of 100 bp
*      51917      52755: contig of 839 bp in length
*      52756      52855: gap of 100 bp
*      52856      53707: contig of 852 bp in length
*      53708      53807: gap of 100 bp
*      53808      54687: contig of 880 bp in length
*      54688      54787: gap of 100 bp
*      54788      55651: contig of 864 bp in length
*      55652      55751: gap of 100 bp
*      55752      56608: contig of 857 bp in length
*      56609      56708: gap of 100 bp
*      56709      57576: contig of 868 bp in length
*      57577      57676: gap of 100 bp
*      57677      58531: contig of 855 bp in length
*      58532      58631: gap of 100 bp
*      58632      59474: contig of 843 bp in length
*      59475      59574: gap of 100 bp
*      59575      60408: contig of 834 bp in length
*      60409      60508: gap of 100 bp
*      60509      61396: contig of 888 bp in length
*      61397      61496: gap of 100 bp
*      61497      62325: contig of 829 bp in length
*      62326      62425: gap of 100 bp
*      62426      63255: contig of 830 bp in length
*      63256      63355: gap of 100 bp
*      63356      64201: contig of 846 bp in length
*      64202      64301: gap of 100 bp
*      64302      65155: contig of 854 bp in length
*      65156      65255: gap of 100 bp
*      65256      66110: contig of 855 bp in length
*      66111      66210: gap of 100 bp
*      66211      67072: contig of 862 bp in length
*      67073      67172: gap of 100 bp
*      67173      68036: contig of 864 bp in length
*      68037      68136: gap of 100 bp
*      68137      69021: contig of 885 bp in length
*      69022      69121: gap of 100 bp
*      69122      69982: contig of 861 bp in length
```

Query Match 12.4%; Score 202.4; DB 2; Length 162548;
Best Local Similarity 71.8%; Pred. No. 2; Se-35;
Matches 293; Conservative 0; Mismatches 108; Indels 7; Gaps 2;

```
OY 1186 CTTTCAGTACCAAGAAATCCAGGCGCTGGAATGTTTACTACGTTAATTTCTAAC 1245
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 153495 CTTTACTGCTANAGNAAACACCCAGACACTTGACACTGCTTTCCAGTTATTTCTAAC 153554

OY 1246 TAAAGAACCTCTGTTGCCAAAAGTATAAAGACAGCCCTGTGAGCTGTGGCAGAGCT 1305
      || ||| ||||| || | | | | | | | | | | | | | | | | | | | | | |
Db 153555 TATGGAATCTGCTGCTGTAGAGAGTGACAGACGACTTGTATTTCCGGCAGAGTGT 153614

OY 1306 GTGACCCCATGTCATTCATTGGGCTCTACTATTAGGAAAGAACAAACACAC 1365
      | |||| | | | | | | | | | | | | | | | | | | | | | | | | | |
```



```

* 20980 22248: contig of 1269 bp in length
* 22249 22348: gap of unknown length
* 22609 23609: contig of 1261 bp in length
* 23610 23709: gap of unknown length
* 23710 24963: contig of 1254 bp in length
* 24964 25063: gap of unknown length
* 25064 26600: contig of 1537 bp in length
* 26601 26700: gap of unknown length
* 26701 27731: contig of 1031 bp in length
* 27732 27831: gap of unknown length
* 27832 29058: contig of 1227 bp in length
* 29059 29159: gap of unknown length
* 29159 30234: contig of 1076 bp in length
* 30235 30334: gap of unknown length
* 30335 31678: contig of 1344 bp in length
* 31679 31778: gap of unknown length
* 31779 33378: contig of 1600 bp in length
* 33379 33478: gap of unknown length
* 33479 34714: contig of 1236 bp in length
* 34715 34814: gap of unknown length
* 34815 36179: contig of 1365 bp in length
* 36180 36279: gap of unknown length
* 36280 37717: contig of 1438 bp in length
* 37718 37817: gap of unknown length
* 37818 39086: contig of 1269 bp in length
* 39087 39186: gap of unknown length
* 39187 40741: contig of 1555 bp in length
* 40742 40841: gap of unknown length
* 40842 42312: contig of 1471 bp in length
* 42313 42412: gap of unknown length
* 42413 43475: contig of 1063 bp in length
* 43476 43575: gap of unknown length
* 43576 45143: contig of 1568 bp in length
* 45144 45243: gap of unknown length
* 45244 46483: contig of 1240 bp in length
* 46484 46583: gap of unknown length
* 46584 48590: contig of 2007 bp in length
* 48591 48690: gap of unknown length
* 48691 50392: contig of 1702 bp in length
* 50393 50492: gap of unknown length
* 50493 52223: contig of 1730 bp in length
* 52223 52322: gap of unknown length
* 52323 53891: contig of 1369 bp in length
* 53892 53991: gap of unknown length
* 53992 55698: contig of 1707 bp in length
* 55699 55798: gap of unknown length
* 55799 57953: contig of 2155 bp in length
* 57954 58053: gap of unknown length
* 58054 59390: contig of 1337 bp in length
* 59391 59490: gap of unknown length
* 59491 61859: contig of 2369 bp in length
* 61860 61959: gap of unknown length
* 61960 63467: contig of 1508 bp in length
* 63468 63567: gap of unknown length
* 63568 65055: contig of 1488 bp in length
* 65056 65155: gap of unknown length
* 65156 66516: contig of 1361 bp in length
* 66517 66616: gap of unknown length
* 66617 68625: contig of 2010 bp in length
* 68626 68726: gap of unknown length
* 68727 70038: contig of 1312 bp in length
* 70039 70138: gap of unknown length
* 70139 71379: contig of 1241 bp in length
* 71380 71479: gap of unknown length
* 71480 73457: contig of 1978 bp in length
* 73458 73557: gap of unknown length
* 73558 75747: contig of 2190 bp in length
* 75748 75847: gap of unknown length
* 75848 77681: contig of 1834 bp in length
* 77682 77781: gap of unknown length
* 77782 79051: contig of 1270 bp in length
* 79052 79151: gap of unknown length
* 79152 81579: contig of 2428 bp in length

```

```

* 81580 81679: gap of unknown length
* 81680 84013: contig of 2334 bp in length

Query Match
Best Local Similarity 69.7%; Pred. No. 1,1e-25;
Matches 301; Conservative 0; Mismatches 118; Indels 13; Gaps 7;

QY 1170 TTGGCTCAGCTTGTGCTTCAATGACCAAG-AAAATCCAGGCGCTTGGATTGTTA 1228
Db 132576 TTGGGCGCACCTGTCTCTTCCCTAGGTCACAAAGAAAGCCAGAACCTGGCAGCTCTT 132635

QY 1229 CTCAGTTAATTTCTACTAAGAACCTGTGTCACAAAGGTATAAACAGAGCCCTTG 1288
Db 132636 GCCAGTTAATGCTAACCAGGAAATGCTGTCTGCCAAAGGT-GCAACAGAGACCTTG 132694

QY 1289 TAGCTGTGGCAGCAGCTGTGACCCCG-ATGCAATCATTTGGGCTCTCACTATTAGG 1347
Db 132695 TATTCCAGGCACAGGTCGACCCCAATGTCATATTTCTCTCACTACACAGAG 132754

QY 1348 ---AAAGAACAAACACCTCACAGCCTTGAAAAAGAAAACACTGTGTCAAAAG--- 1401
Db 132755 GAAAAACTAACAAACACAGCCTCATGTGTGAAAAAGTAACCTATATGCCAAATGGGAA 132814

QY 1402 GAAAAATATTCACCCCATTAATAATTAAGAAACAGACAGATCATTTGGAG-- 1459
Db 132815 GAAAAATTTTACCCCAACAAACATCTCAAGGGCGAAGACAGATTAATTGGAGGG 132874

QY 1460 -GAGAGATTGCCAGTGGGGGACAGATGTATATATAGATATGAAGTCACCTACTTGA 1518
Db 132875 AAGAGGCCAGCCAAAGGCGACATATATATATATATATATATATATATATATATAT 132934

QY 1519 AAAGATTAAATTTACTT-ATTGCTTTCAGGTAAGGCTATATGACCTCTCACTCTCC 1577
Db 132935 TGTGAATGACACAGCTCTGTCTGTGTTGAGTAAGACTGTCTGAGCTGTGATCTCTCG 132994

QY 1578 TAGCCACTCTCTC 1589
Db 132995 TATCCACTCTCTC 133006

RESULT 8
AC125147 219386 bp DNA linear HTG 22-JUN-2002
LOCUS Mus musculus chromosome UNK clone R24-350A15, WORKING DRAFT
DEFINITION
AC125147
VERSION AC125147.2 GI:21539180
ACCESSION AC125147
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 219386)
AUTHORS McKernson,J.D. and Waterston,R.H.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 219386)
AUTHORS McKernson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 219386)
AUTHORS McKernson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (22-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
On Jun 22, 2002 this sequence version replaced gi:21490647.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu

```


Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,
Williams, G., Williamson, A., Wleczek, R., Wooden, S., Morley, K.,
Wu, C., Wu, Y., Wu, Y., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 82394)
Morley, K.C.
Submitted (15-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 82394)
Morley, K.C.
Direct Submission
Submitted (17-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 14, 2002 this sequence version replaced gl:20258108.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GNGF
Center clone name: CH230-366D18
----- Summary Statistics
Sequencing vector: Plasmid:
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 2214 bases at least Q40
Consensus quality: 24677 bases at least Q30
Consensus quality: 26339 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_drafl_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 51 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1022: contig of 1022 bp in length
1023 1122: gap of unknown length
1123 2161: contig of 1039 bp in length
2162 2261: gap of unknown length
2262 3382: contig of 1121 bp in length
3383 3482: gap of unknown length
3483 4532: contig of 1050 bp in length
4533 4632: gap of unknown length
4633 5738: contig of 1106 bp in length
5739 5838: gap of unknown length
5839 6855: contig of 1017 bp in length
6856 6956: gap of unknown length
6957 8043: contig of 1068 bp in length
8044 8143: gap of unknown length
8144 9165: contig of 1022 bp in length
9166 9265: gap of unknown length
9266 10660: contig of 1395 bp in length
10661 10760: gap of unknown length
10761 11876: contig of 1116 bp in length
11877 11976: gap of unknown length
11978 13290: contig of 1314 bp in length
13291 13390: gap of unknown length
13391 14419: contig of 1029 bp in length
14420 14519: gap of unknown length
14520 15853: contig of 1336 bp in length
15854 15955: gap of unknown length
15956 17092: contig of 1137 bp in length
17093 17192: gap of unknown length

17193 18678: contig of 1486 bp in length
18679 18778: gap of unknown length
18779 19963: contig of 1185 bp in length
19964 20063: gap of unknown length
20064 21418: contig of 1355 bp in length
21419 21518: gap of unknown length
21519 22632: contig of 1114 bp in length
22633 22732: gap of unknown length
22733 24482: contig of 1750 bp in length
24483 24582: gap of unknown length
24583 26077: contig of 1495 bp in length
26078 26177: gap of unknown length
26178 27428: contig of 1251 bp in length
27429 27528: gap of unknown length
27529 28784: contig of 1256 bp in length
28785 28884: gap of unknown length
28885 30527: contig of 1643 bp in length
30528 30627: gap of unknown length
30628 32098: contig of 1471 bp in length
32099 32198: gap of unknown length
32199 33420: contig of 1222 bp in length
33421 33520: gap of unknown length
33521 34996: contig of 1476 bp in length
34997 35096: gap of unknown length
35097 36923: contig of 1827 bp in length
36924 37023: gap of unknown length
37024 38600: contig of 1837 bp in length
38601 38960: gap of unknown length
38961 40176: contig of 1216 bp in length
40177 40276: gap of unknown length
40277 42077: contig of 1801 bp in length
42078 42177: gap of unknown length
42178 43618: contig of 1441 bp in length
43619 43718: gap of unknown length
43719 44769: contig of 1051 bp in length
44770 44869: gap of unknown length
44871 45887: contig of 1018 bp in length
45888 45987: gap of unknown length
45989 48401: contig of 2414 bp in length
48402 48501: gap of unknown length
48502 49992: contig of 1491 bp in length
49993 50092: gap of unknown length
50093 52227: contig of 2135 bp in length
52228 52327: gap of unknown length
52328 53979: contig of 1652 bp in length
53980 54079: gap of unknown length
54080 55974: contig of 1895 bp in length
55975 56074: gap of unknown length
56075 57610: contig of 1436 bp in length
57611 59661: contig of 1951 bp in length
59662 61354: contig of 1693 bp in length
61355 61454: gap of unknown length
61455 63275: contig of 1821 bp in length
63276 63375: gap of unknown length
63376 64961: contig of 1586 bp in length
64962 65061: gap of unknown length
65062 66893: contig of 1832 bp in length
66894 66993: gap of unknown length
66994 68675: contig of 1582 bp in length
68676 68776: gap of unknown length
68777 70714: contig of 2039 bp in length
70715 70814: gap of unknown length
70815 72329: contig of 1515 bp in length
72330 74582: gap of unknown length
74583 74682: contig of 2153 bp in length
74683 76995: gap of unknown length
76996 77095: contig of 2313 bp in length
77096 79610: contig of 2515 bp in length
79611 82394: contig of 2684 bp in length


```

FEATURES
  source
    location="/20110101"
    1. 259782
    /organism="Rattus norvegicus"
    /db_xref="taxon:10116"
    /clone="CH230-2D5"

```

Work on the sequencing that the 5' ends of the contigs were based on the understanding that the 3' ends of the contigs may be contaminated with foreign DNA. The sequence may be contaminated with foreign DNA. continue. The sequence may be contaminated with foreign DNA. from E. coli, yeast, vector, phage, etc.

contigs composition:

16662 bp contig from 1 to 16662

16662 bp contig from 16763 to 23333

6571 bp contig from 23434 to 79741

56308 bp contig from 79842 to 111924

32083 bp contig from 112025 to 120635

8612 bp contig from 120737 to 131685

10929 bp contig from working draft sequence. It currently consists of 6 contigs. Gaps between the pieces

NOTE: This is a contig. Gaps between the pieces are represented as runs of N. The order of the sizes are represented as correct as given, however the sizes are believed to be correct as given, based on estimates that have

of the gaps between the submitters.

provided by the submitter.

This sequence will be replaced

by the finished sequence as soon as it is available and

the accession number will be preserved.

16662 bp in length

16662 bp

100 bp

16663 16762: gap of


```
Project Information
Project Name: 0
Center
Center clone name: RPCI-23_197M9
```

[illegible]

34646	34745:	gap of unknown length
34746	35894:	contig of 1149 bp in length
35895	35994:	gap of unknown length
35995	37156:	contig of 1162 bp in length
37157	37356:	gap of unknown length
37357	38396:	contig of 1140 bp in length
38397	38496:	gap of unknown length
38497	39570:	contig of 1074 bp in length
39571	39670:	gap of 1182 bp in length
39671	40852:	contig of unknown length
40853	40952:	gap of 1164 bp in length
40953	42116:	contig of unknown length
42117	42216:	gap of unknown length
42217	43372:	contig of 1156 bp in length
43373	44472:	gap of unknown length
44473	44603:	contig of 1131 bp in length
44603	44703:	gap of unknown length
44704	45924:	contig of 1221 bp in length
45925	46024:	gap of unknown length
46025	47174:	contig of 1150 bp in length
47175	47274:	gap of unknown length
47275	48744:	contig of 1470 bp in length
48745	48844:	gap of unknown length
48845	49699:	contig of 1125 bp in length
49699	50069:	gap of unknown length
50070	51135:	contig of 1066 bp in length
51136	51230:	gap of unknown length
51236	52590:	contig of 1455 bp in length
52591	52790:	gap of unknown length
52791	53291:	contig of 1130 bp in length
53291	54020:	gap of unknown length
54021	55101:	contig of 1081 bp in length
55102	55201:	gap of unknown length
55202	56344:	contig of 1123 bp in length
56345	56424:	gap of unknown length
56425	57583:	contig of 1159 bp in length
57584	57683:	gap of unknown length
57684	58822:	contig of 1139 bp in length
58823	58922:	gap of unknown length
58923	60547:	contig of 1625 bp in length
60548	60647:	gap of unknown length
60649	61743:	contig of 1096 bp in length
61744	61843:	gap of unknown length
61844	62998:	contig of 1155 bp in length
62999	63098:	gap of unknown length
63099	64220:	contig of 1122 bp in length
64221	64330:	gap of unknown length
64331	65388:	contig of 1078 bp in length
65389	65498:	gap of unknown length
65499	66762:	contig of 1264 bp in length
66763	66862:	gap of unknown length
66863	66823:	contig of 1153 bp in length
66824	68025:	gap of unknown length
68026	68125:	gap of 1275 bp in length
68126	69400:	contig of unknown length
69401	69500:	gap of unknown length
69501	70718:	contig of 1218 bp in length
70719	70818:	gap of unknown length
70819	71995:	contig of 1177 bp in length
71996	72095:	gap of unknown length
72096	73232:	contig of 1137 bp in length
73233	73332:	gap of unknown length
73333	74476:	contig of 1144 bp in length
74477	74576:	gap of unknown length
74577	75863:	contig of 1287 bp in length
75864	75963:	gap of unknown length
75964	77021:	contig of 1058 length
77022	77121:	gap of unknown length
77122	78312:	contig of 1091 bp in length
78313	78412:	gap of unknown length
78414	79444:	contig of 1112 bp in length
79445	79454:	gap of unknown length
79455	80617:	contig of 1093 bp in length
80618	80717:	gap of unknown length

us-09-605-042a-37.rge

Wed Feb 19 14:09:20 2003

```
*      80718      81804: contig of 1087 bp in length
*      81805      81904: gap of unknown length
*      83007      83007: contig of 1103 bp in length
*      83108      83107: gap of unknown length
*      83108      84167: contig of 1060 bp in length
*      84168      84267: gap of unknown length
*      84268      85347: contig of 1080 bp in length
*      85348      85447: gap of unknown length
*      85448      85348: contig of 1074 bp in length
*      86522      86521: gap of unknown length
*      86522      86522: contig of 1123 bp in length
*      87745      87744: gap of unknown length
*      87845      87845: contig of 1121 bp in length
*      88966      88965: gap of unknown length
*      89066      89065: gap of unknown length
*      90177      90176: contig of 1111 bp in length
*      90277      90276: gap of unknown length
*      91280      91279: gap of 1003 bp in length
*      91280      91279: contig of unknown length
*      91379      91379: gap of 1184 bp in length
*      92563      92563: contig of unknown length
*      92563      92563: gap of 1025 bp in length
*      92564      92564: contig of unknown length
*      92564      92564: gap of unknown length
*      93689      93688: contig of unknown length
*      93789      93788: gap of 1146 bp in length
*      94935      94934: contig of unknown length
*      95035      95034: gap of unknown length
*      96112      96111: contig of 1077 bp in length
*      96211      96211: gap of unknown length
*      96211      96211: contig of 2141 bp in length
*      98352      98352: gap of unknown length
*      98453      98453: contig of 1193 bp in length
*      98453      98453: gap of unknown length
*      99745      99745: gap of unknown length
*      99646      99646: gap of unknown length

Query Match      3.2%; Score 51.4; DB 2; Length 303091;
Best Local Similarity 52.0%; Pred No. 0.05; Indels 0; Gaps 0;
Matches 115; Conservative 0; Mismatches 106;

QY 111 AGAAAAAGATTTAGAACTTTATTTCTTCTGAGATTTATGCAAGAAAGAGGAAAA 170
Db 12213 AAAAAAAAAAAAAAAAAATAGTGAATAATTAATAATTAATAAAAAAAAAAAAA 12272
QY 171 AAAAAACATTTCTTATGAGGATTAACGGGCAAGATACAAACGTTCCAGAAAAAGATA 230
Db 12273 GAAAAAAAAATATATAAAAAAAAAAGAAAAAAGAAAAATATAAAAAAAAAAAAAATA 12332
QY 231 AATAGTACAAATGAAAAAGTAACTCTTTCTCAACAAGAACCGCAAAAGTAAATA 12392
Db 12333 TAAAAAAAAAAAAAAAAAGTAAAAAAGAAAAAAGAAAAAATATAATAATA 12392
QY 291 TGAATAGATGTTCTCACTTTCCCAAGAGATGAAGTAA 331
Db 12393 AAAAAAAAAATTAATAAAAAAAAAAAAAAAAAAGAAAAATGA 12433

Search completed: February 17, 2003, 15:35:06
Job time : 5250.6 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 17, 2003, 08:54:07 ; Search time 67.3488 Seconds
(Without alignments)
9111.676 Million cell updates/sec

Title: US-09-605-042a-1_COPY_4000_6000

Perfect score: 2001
1 aagaactcttgataaca.....tgtatatatacatataca 2001

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 segs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_NA:*
1: /cgn2_6/ptodata/2/1na/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/1na/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/1na/5A_COMB.seq:*
4: /cgn2_6/ptodata/2/1na/5B_COMB.seq:*
5: /cgn2_6/ptodata/2/1na/PTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/1na/Backlist1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	86	4.3	7218	1	US-08-232-463-14
2	59.2	3.0	6727	3	US-08-629-643A-5
3	59.2	3.0	6727	3	US-09-280-799-1
4	59.2	3.0	6727	4	US-09-155-884-5
5	58.2	2.9	3288	2	US-08-620-94A-1
6	58.2	2.9	3288	3	US-09-022-255-1
7	58.2	2.9	3288	3	US-09-022-696-1
8	58.2	2.9	3288	3	US-08-978-773-1
9	58.2	2.9	3288	3	US-09-022-253-1
10	58.2	2.9	3288	3	US-09-022-260-1
11	58.2	2.9	3288	4	US-09-022-259-1
12	58.2	2.9	3288	4	US-09-022-257-1
13	58	2.9	4223	4	US-08-845-258-7
14	58	2.9	4223	4	US-08-845-258-7
15	58	2.9	4223	4	US-08-990-571-7
16	58	2.9	4223	4	US-08-990-571-7
17	58	2.9	4223	4	US-08-990-571-7
18	58	2.9	4223	4	US-08-723-142A-7
19	58	2.9	4223	4	US-08-723-142A-7
20	58	2.9	4223	4	US-08-528-784A-7
21	55.4	2.8	4223	4	US-09-528-784A-45
22	55.4	2.8	3381	4	US-09-371-507-1
23	55.4	2.8	3381	4	US-09-371-507-1
24	55.2	2.8	4550	4	US-09-103-663-35
25	53.6	2.7	45546	4	US-09-146-053-6
26	53.4	2.7	4771	2	US-08-866-650-2
27	53.4	2.7	4771	2	US-09-021-287-2

C 28	53.4	2.7	4771	4	US-09-240-473-2	Sequence 2, Appli
29	53.2	2.7	3358	3	US-08-957-063-4	Sequence 4, Appli
30	53.2	2.7	3358	4	US-09-487-685-4	Sequence 4, Appli
31	53.2	2.7	3358	4	US-08-802-805D-4	Sequence 4, Appli
32	52.8	2.6	870	2	US-08-420-629-9	Sequence 9, Appli
33	52.8	2.6	2551	1	US-08-486-013-70	Sequence 70, Appli
34	52.8	2.6	2551	2	US-08-486-013-70	Sequence 70, Appli
35	52.8	2.6	2551	2	US-08-482-279-70	Sequence 70, Appli
36	52.8	2.6	2551	3	US-08-342-268-70	Sequence 70, Appli
37	52.8	2.6	2551	4	US-09-015-968-70	Sequence 70, Appli
38	52.8	2.6	4718	3	US-09-397-386-70	Sequence 70, Appli
39	52.8	2.6	4733	3	US-08-936-135-11	Sequence 9, Appli
40	52.8	2.6	4769	3	US-08-936-135-13	Sequence 13, Appli
41	52.8	2.6	4784	3	US-08-936-135-15	Sequence 15, Appli
42	52.6	2.6	884	4	US-09-171-209-76	Sequence 76, Appli
43	52.4	2.6	2997	3	US-08-486-343A-3	Sequence 3, Appli
44	52.4	2.6	2997	5	PCR-US95-07349-3	Sequence 3, Appli
45	52.4	2.6	80595	4	US-09-078-294-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PTZ9PL-F15
; US-08-232-463-14
Query Match 4.3%; Score 86; DB 1; Length 7218.

Best Local Similarity 5.5%; Pred. No. 2.2e-15;
Matches 23; Conservative 249; Mismatches 144; Indels 0; Gaps 0;

RESULT 2
 US-08-629-643A-5/c
 : Sequence 5, Application US/08629643A
 Patent No. 6025539
 : GENERAL INFORMATION:
 APPLICANT: Lee, J. L.
 APPLICANT: Lee, N. A.
 TITLE OF INVENTION: IL-5 TRANSGENIC MOUSE
 NUMBER OF SEQUENCES: 5
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Schweigman, Lundberg, Woessner & Kluth, P. A.
 STREET: P. O. Box 2938
 CITY: Minneapolis
 STATE: MN
 COUNTRY: USA
 ZIP: 55402
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/629,643A
 FILING DATE:
 CLASSIFICATION: 800
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/629,643
 FILING DATE: 09-APR-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Visknins, Ann S
 REGISTRATION NUMBER: 37,748
 REFERENCE/DOCKET NUMBER: 150.167W01
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 612-359-3260
 TELEFAX: 612-359-3263
 TELEX:
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 6727 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single

```

?      TOPOLOGY:      linear
?      MOLECULE TYPE: Genomic DNA
?
?      FEATURE:
?
?      NAME/KEY:      mat_peptide
?      LOCATION:      650..3771
?      OTHER INFORMATION: Join 650..730, 1560..1592
?
?      OTHER INFORMATION: 3468..3556, 3676..3771
US-08-629-643A-5

```

Query Match	3.0%;	Score	59.2;	DB	3;	Length	6727;
Best local Similarity	83.8%;	Pred.	No. 2.5e-07;				
Matches	67;	Conservative	0;	Mismatches	13;	Indels	0;
						Caps	0

Oy	1922	CATTATTCACAAACACACACATATTTTACACATACACACACACAGCTTTAATA	1984
Db	2653	CATCATCATCATCATATATATATATATACACACACACACACACATATATA	2594
Oy	1982	TGTTATATATATATACATACA	2001
Db	2593	TATATATATATATATATATA	2574

RESULT 3
US-09-280-799-1/c
; Sequence 1, Application US/09280799

```

; GENERAL INFORMATION:
; APPLICANT: Dean, Nicholas M.

```

```

: APPLICANT: Karras, James G
: APPLICANT: McKay, Robert
: TITLE OF INVENTION: ANTISENSE MODULATION OF INTERLEUKIN-5 SIGNAL
: TITLE OF INVENTION: TRANSDUCTION
: FILE REFERENCE: ISPH-0340
: CURRENT APPLICATION NUMBER: US/09/280,799
: CURRENT FILING DATE: 1999-03-26
: NUMBER OF SEQ ID NOS: 208
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO: 1
: LENGTH: 6727
: TYPE: DNA
: ORGANISM: Mus musculus
: US-09-280-799-1

```

Query Match	3.0%	Score	59.2	DB	3	Length	6727
Best Local Similarity	83.8%	Pred	No. 2.5e-07				
Matches	67	Conservative	0	Mismatches	13	Indels	0
						Gaps	0

QY	1982	TGTATATATATATACATACA	2001
QY	1922	CAATTAATCAAAACACACACACATATATATACACATATACACACACACACACACACATATATATA	1981
Db	2653	CATACATATACATCATATCATATATATATATATATACACACACACACACACACACACATATATA	2594
QY	1982	TGTATATATATATATACATACA	2001
Db	2593	TATATATATATATATATATATA	2574

RESULT 4
US-09-155-884-5/c
; Sequence 5, Application US/09155884

```

; GENERAL INFORMATION:
;
; APPLICANT: James J. Lee et al.
;

```

```

; TITLE OF INVENTION: IL-5 TRANSGENIC MOUSE
;
; NUMBER OF SEQUENCES: 5
;
; CORRESPONDENCE ADDRESS:
;

```

ADDRESS: Schwegman, Lundberg, Woessner & Kluth, P. A.
STREET: P. O. Box 2938
CITY: Minneapolis

```

; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
;

```

FILING DATE: 7 AUGUST 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA: USSN 08/410,535
APPLICATION NUMBER: 23 MARCH 1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Peirkins, Patricia Anne
REGISTRATION NUMBER: 34,695-B
REFERENCE/DOCKET INFORMATION:
TELEPHONE: (206)587-0430
TELEFAX: (206) 1:
INFORMATION FOR SEQ. ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3288 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: HVS13 receptor
STRAIN:
FEATURE:
NAME/KEY: CDS
LOCATION: 121..2715
US-08-620-694A-1

Query Match 2.9%; Score 58.2; DB 2; Length 3288;
Best Local Similarity 63.8%; Pred No. 3.2e-07; Indels 1; Gaps 1
Matches 104; Conservative 0; Mismatches 58;

Db 1839 GATCCTTATATCTACACCTCTCAGTCCAGATTATAGTGTGCACCACTTATCTACAGTT 1958
2922 GATCATCAGACGTGTACAGTCCAGATGCTGGGACTTCAGGTATACACTAT-GTCAAGT 1958
GATCATCAGACGTGTGTAAAGCGGGGCACTACACACACACACACATATATACACATAC 2804
Gy 1899 TATGCTGTGTAGAGCTACAGCCATTATTAACAACACACACATATATACACATAC 1958
2863 TCAGATGTCACATTTCTTAAGCGGGGCACTACACACACACACACACACACACAC 2761
Db 1959 ACACGACACACACAGTATATATATATATATATATATACATACA 2001
2803 ACACGACACACACACACACACACACACACACACACACACGACGACGACAC 2761
Gy 2803

RESULT 6
US-09-022-255-1/c Application US/09022255
Sequence 1, Application US/09022255
Patent No. 6072033
GENERAL INFORMATION:
APPLICANT: Yao, Zhengbin
APPLICANT: Spriggs, William
APPLICANT: Fanslow, No. 6072033el Receptor That Binds IL-17
TITLE OF INVENTION: 10
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Power Macintosh
OPERATING SYSTEM: Apple Operating System 7.5.5
SOFTWARE: Microsoft Word for Apple, Version 6.0.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/022,255
FILING DATE:
CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: USSN 08/620,694
FILING DATE: 21 MARCH 1996
APPLICATION NUMBER: USSN 08/538,765
FILING DATE: 7 AUGUST 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: USSN 08/410,535
FILING DATE: 23 MARCH 1995
ATTORNEY/AGENT INFORMATION:
NAME: Perkins, Patricia Anne
REGISTRATION NUMBER: 34,695
REFERENCE/DOCKET NUMBER: 2617-B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)587-0430
TELEFAX: (206)
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3288 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Mouse
STRAIN: HVS13 receptor
FEATURE:
NAME/KEY: CDS
LOCATION: 121..2715
US-09-022-255-1

Query Match 2.9%; Score 58.2; DB 3; Length 3288;
Best Local Similarity 63.8%; Pred. No. 3.2e-07;
Matches 104; Conservative 0; Mismatches 58; Indels 1; Gaps 1;
QY 1839 GATCCCTAATCTACACTCCAGAGGCCAGATTTAAGTGGACACATATCTAGTT 1898
DB 2922 GATCATCAAGTCTAGTCCCAAGTCTGGGACTTCAGGTATACACACTAT-GTCA GAT 2864
QY 1899 TATGCTGTCTAAGAGTAAAGCCCAATTATACAAACACACACATATATACACATAC 1958
DB 2863 TCAGATTTTCACTTTTCTAAGCCGGGCACTACACACACACACACACACACACAC 2804
QY 1959 ACACACACACACACGATATATATATATATATATATATATATATATATATATATAC 2001
DB 2803 ACACGCAC 2761

RESULT 7
US-09-022-696-1/c
Sequence 1, Application US/09022696
Patent No. 6072037
GENERAL INFORMATION:
APPLICANT: Yao, Zhenbin
APPLICANT: Spriggs, Melanie
APPLICANT: Fanslow, William
TITLE OF INVENTION: No. 6072037e1 Receptor That Binds IL-17
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Power Macintosh
OPERATING SYSTEM: Apple Operating System 7.5.5
SOFTWARE: Microsoft Word for Apple, Version 6.0.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/022,696

FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/620,694
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: USSN 08/410,535
FILING DATE: 23 MARCH 1995
ATTORNEY/AGENT INFORMATION:
NAME: Perkins, Patricia Anne
REGISTRATION NUMBER: 34,695
REFERENCE/DOCKET NUMBER: 2617-B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)587-0430
TELEFAX: (206)
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3288 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Mouse
STRAIN: HVS13 receptor
FEATURE:
NAME/KEY: CDS
LOCATION: 121..2715
US-09-022-696-1

Query Match 2.9%; Score 58.2; DB 3; Length 3288;
Best Local Similarity 63.8%; Pred. No. 3.2e-07;
Matches 104; Conservative 0; Mismatches 58; Indels 1; Gaps 1;
QY 1839 GATCCCTAATCTACACTCCAGAGGCCAGATTTAAGTGGACACATATCTAGTT 1898
DB 2922 GATCATCAAGTCTAGTCCCAAGTCTGGGACTTCAGGTATACACACTAT-GTCA GAT 2864
QY 1899 TATGCTGTCTAAGAGTAAAGCCCAATTATACAAACACACATATATACACATAC 1958
DB 2863 TCAGATTTTCACTTTTCTAAGCCGGGCACTACACACACACACACACACACACAC 2804
QY 1959 ACACACACACACACGATATATATATATATATATATATATATATATATATATATAC 2001
DB 2803 ACACGCAC 2761

RESULT 8
US-08-978-773-1/c
Sequence 1, Application US/08978773
Patent No. 6083906
GENERAL INFORMATION:
APPLICANT: Trout, Anthony
TITLE OF INVENTION: Method of Regulating Nitric Oxide Production
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple PowerMacintosh
OPERATING SYSTEM: Apple Operating System 7.5.5
SOFTWARE: Microsoft Word for PowerMacintosh, Version 6.0.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/978,773

Wed Feb 19 14:09:10 2003

us-09-605-042a-1_copy-4000-6000.rni

FILING DATE: 530
CLASSIFICATION: 530
PRIOR APPLICATION DATA: USSN 60/052,525
APPLICATION NUMBER: 27 NOVEMBER 1996
FILING DATE: 27 NOVEMBER 1996
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Perkins, Patricia Anne
REGISTRATION NUMBER: 34,693
REFERENCE/DOCKET NUMBER: 2623-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)587-0430
TELEFAX: (206)587-0430
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3288 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE: MOUSE
ORGANISM: MOUSE
IMMEDIATE SOURCE: IL-17 receptor
CLONE: IL-17 receptor
FEATURE:
NAME/KEY: CDS
LOCATION: 121..2712
US-09-605-042a-1

Query Match 2.9%; Score 58.2; DB 3; Length 3288; Indels 1; Gaps 1;
Best Local Similarity 63.8%; Pred. No. 3.2e-07; Mismatches 58;
Matches 104; Conservative 0;
DB 1839 GATCCTTATCTACACTCTCAAGTCCCGCATTTAAGTGTGCACACCTATCTACTAGT 1898
DB 2922 GATCCTTATCTACACTCTCAAGTCCCGCATTTAAGTGTGCACACCTATCTACTAGT 1898
DB 1839 TATGCTGTCTAAGACCTTACAGCCGATTTATACACACACACACACACACACACAC 1958
DB 1839 TATGCTGTCTAAGACCTTACAGCCGATTTATATATATATATATATATATATATAT 1958
DB 2863 TCAGATGTTTCACATTTCTAAGCCGGGACCTACACACACACACACACACACACAC 2804
DB 1959 ACACACACACACACGATATATATATATATATATATATATATATATATATATAT 2804
DB 2803 ACACGAC 2761

CLASSIFICATION: 530
PRIOR APPLICATION DATA: USSN 08/620,694
APPLICATION NUMBER: 21-MARCH-1996
FILING DATE: 21-MARCH-1996
APPLICATION NUMBER: USSN 08/538,765
FILING DATE: 7 AUGUST 1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA: USSN 08/410,535
APPLICATION NUMBER: 23 MARCH 1995
FILING DATE: 23 MARCH 1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Perkins, Patricia Anne
REGISTRATION NUMBER: 34,695
REFERENCE/DOCKET NUMBER: 2617-B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)587-0430
TELEFAX: (206)587-0430
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3288 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE: MOUSE
ORGANISM: MOUSE
IMMEDIATE SOURCE: HVS13 receptor
CLONE: HVS13 receptor
FEATURE:
NAME/KEY: CDS
LOCATION: 121..2715
US-09-605-042a-1

Query Match 2.9%; Score 58.2; DB 3; Length 3288; Indels 1; Gaps 1;
Best Local Similarity 63.8%; Pred. No. 3.2e-07; Mismatches 58;
Matches 104; Conservative 0;
DB 1839 GATCCTTATCTACACTCTCAAGTCCCGCATTTAAGTGTGCACACCTATCTACTAGT 1898
DB 2922 GATCCTTATCTACACTCTCAAGTCCCGCATTTAAGTGTGCACACCTATCTACTAGT 1898
DB 1839 TATGCTGTCTAAGACCTTACAGCCGATTTATACACACACACACACACACACACAC 1958
DB 1839 TATGCTGTCTAAGACCTTACAGCCGATTTATATATATATATATATATATATATAT 1958
DB 2863 TCAGATGTTTCACATTTCTAAGCCGGGACCTACACACACACACACACACACACAC 2804
DB 1959 ACACACACACACACGATATATATATATATATATATATATATATATATATATAT 2804
DB 2803 ACACGAC 2761

RESULT 9
US-09-022-253-1/c Application us/09022253
Sequence 1, Application us/09022253
Patent No. 6096305
GENERAL INFORMATION:
APPLICANT: Yao, Zhengbin
APPLICANT: Spriggs, William
APPLICANT: Farnlow, William
TITLE OF INVENTION: No. 6096305el Receptor That Binds IL-17
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Power Macintosh
OPERATING SYSTEM: Apple Operating System 7.5.5
SOFTWARE: Microsoft Word for Apple, version 6.0.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: us/09/022,253
FILING DATE.

RESULT 10
US-09-022-260-1/c Application us/09022260
Sequence 1, Application us/09022260
Patent No. 6100235
GENERAL INFORMATION:
APPLICANT: Yao, Zhengbin
APPLICANT: Spriggs, William
APPLICANT: Farnlow, William
TITLE OF INVENTION: No. 6100235el Receptor That Binds IL-17
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Power Macintosh
OPERATING SYSTEM: Apple Operating System 7.5.5

Wed Feb 19 14:09:10 2003

us-09-605-042a-1_copy_4000_6000.rml

```

SOFTWARE: Microsoft Word for Apple, Version 6.0.1
CURRENT APPLICATION DATA: US/09/022,260
APPLICATION NUMBER: 08/620,694
CLASSIFICATION:
FILING DATE:
PRIORITY APPLICATION NUMBER: 23 MARCH 1995
PRIORITY APPLICATION NUMBER: 08/620,694
CLASSIFICATION:
PRIORITY APPLICATION NUMBER: USSN 08/410,535
FILING DATE: 23 MARCH 1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Perkins, Patricia Anne
REGISTRATION NUMBER: 34,695
REFERENCE/DOCKET NUMBER: 2617-B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)587-0430
TELEFAX: (206)
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3288 base pairs
TYPE: nucleic acid
STRANDEDNESS: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTISENSE: NO
ORIGINAL SOURCE:
ORGANISM: Mouse
STRAIN: HVS13 receptor
FEATURE:
NAME/KEY: CDS
LOCATION: 121..2715
US-09-022-260-1
Query Match
Best Local Similarity 63.8%; Score 58.2; DB 3; Length 3288;
Matches 104; Conservative 0; Mismatches 58; Indels 1; Gaps 1;
QY 1839 GATCCTTATCTACAGCTCAAGTCCCAATATATATAGTGTGACACCATATATACATGTT 1898
DB 2922 GATCATCAAGTCTCAGTCCCAAGTCTGCGACTTCAGGTATACACACTAT-GTCAGAT 2864
QY 1899 TATGCTGTCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1958
DB 2863 TCAGATGTTTCACTTTCTTAAGCGGCGGACCTACACACACACACACACACACACAC 2804
QY 1959 ACACACACACACACGATATATATATATATATATATATATATATATATATATATATAC 2001
DB 2803 ACACGACACACACACACACACACACACACACACACACACACACACACACACACACACAC 2761

```

```

COMPUTER: Apple Power Macintosh
OPERATING SYSTEM: Apple Operating System 7.5.5
SOFTWARE: Microsoft Word for Apple, Version 6.0.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/022,259
CLASSIFICATION:
FILING DATE:
PRIORITY APPLICATION NUMBER: 08/620,694
PRIORITY APPLICATION NUMBER: 23 MARCH 1995
PRIORITY APPLICATION NUMBER: USSN 08/410,535
FILING DATE: 23 MARCH 1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Perkins, Patricia Anne
REGISTRATION NUMBER: 34,695
REFERENCE/DOCKET NUMBER: 2617-B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)587-0430
TELEFAX: (206)
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3288 base pairs
TYPE: nucleic acid
STRANDEDNESS: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTISENSE: NO
ORIGINAL SOURCE:
ORGANISM: Mouse
STRAIN: HVS13 receptor
FEATURE:
NAME/KEY: CDS
LOCATION: 121..2715
US-09-022-259-1
Query Match
Best Local Similarity 63.8%; Score 58.2; DB 4; Length 3288;
Matches 104; Conservative 0; Mismatches 58; Indels 1; Gaps 1;
QY 1839 GATCCTTATCTACAGCTCAAGTCCCAATATATATAGTGTGACACCATATATACATGTT 1898
DB 2922 GATCATCAAGTCTCAGTCCCAAGTCTGCGACTTCAGGTATACACACTAT-GTCAGAT 2864
QY 1899 TATGCTGTCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1958
DB 2863 TCAGATGTTTCACTTTCTTAAGCGGCGGACCTACACACACACACACACACACACAC 2804
QY 1959 ACACACACACACACGATATATATATATATATATATATATATATATATATATATATAC 2001
DB 2803 ACACGACACACACACACACACACACACACACACACACACACACACACACACACACACAC 2761

```

```

RESULT 11
US-09-022-259-1/C
Sequence 1, Application US/09022259
Patent No. 6191104
GENERAL INFORMATION:
APPLICANT: Yao, Zhengbin
APPLICANT: Spriggs, Melanie
APPLICANT: Fanslow, William
TITLE OF INVENTION: No. 6191104e1 Receptor That Binds IL-17
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```


COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Power Macintosh
OPERATING SYSTEM: Apple Operating System Version 6.0.1-1
SOFTWARE: Microsoft Word for Apple, Version 6.0.1-1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/022,257

FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA: 08/620,694
 APPLICATION NUMBER:
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA: USN 08/410,535
 APPLICATION NUMBER: 23 MARCH 1995
 FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Perkins, Patricia Anne
 REGISTRATION NUMBER: 34,695
 REFERENCE/DOCKET NUMBER: 2617-B
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206)587-0430
 TELEFAX: (206) ID NO: 1:
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3288 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA to mRNA
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Mouse
 STRAIN: HVS13 receptor
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 121..2715

RESULT 13
 US-08-845-258-7/C
 Sequence 7, Application US/08845258
 Patent No. 6183976
 GENERAL INFORMATION:
 APPLICANT: Reed, Steven G.
 APPLICANT: Lodes, Michael J.
 APPLICANT: Houghton, Raymond
 APPLICANT: Sleeth, Paul R.
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
 AND TREATMENT OF B. MICROIDI INFECTION
 TITLE OF INVENTION: 53
 NUMBER OF SEQUENCES:
 CORRESPONDENCE ADDRESS:
 CORRESPONDENT: SEED AND BERRY
 ADDRESS: 6300 Columbia Center, 701 Fifth Avenue
 STREET

CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
CURRENT APPLICATION NUMBER: US/08/845,258
APPLICATION NUMBER: 24-APR-1997
FILING DATE: 435
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: MAKI, David J. 31,392
REGISTRATION NUMBER: 210121,426C1
REFERENCE/DOCKET INFORMATION:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)682-6031 7:
TELEFAX: FOR SEQ ID NO: 7:
INFORMATION CHARACTERISTICS:
SEQUENCE CHARACTERISTICS:
LENGTH: nucleic acid
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

[illegible]

RESULT 14
US-08-845-258-45
Sequence 45, Application US/08845258
Patent No. 6183976
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond
APPLICANT: Sleeth, Paul R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
TITLE OF INVENTION: 53
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY
ADDRESS: 6300 Columbia Center, 701 Fifth Avenue
STREET: Seattle
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/845,258
FILING DATE: 24-APR-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.426C1

us-09-605-042a-1_copy-4000-6000.rni

Db 3636 ACACACGCACACAGCAGGCACACACGAC 3607
Search completed: February 17, 2003, 16:12:32
Job time : 154.349 secs

TELECOMMUNICATION INFORMATION
TELEPHONE: (206) 632-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 4223 base pairs
type: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-845-258-45	TOPBLAST	Score 58;	DB 4;	Length 4223;	Gaps 0;
Query Match	7.9%;	Pred. No. 4.3e-07;	Indels 0;		
Best Local Similarity	77.8%;	Mismatches 20;			
Matches	70;	Conservative			

[illegible]

RESULT 15
US-08-990-571-7/c
US-08-990-571-7/c Application US/08990571

Sequence 7 Application
Patent No. 6214971
GENERAL INFORMATION: Steven G. et al. METHODS FOR THE DIAGNOSIS AND TREATMENT OF B.
APPLICANT: Reed, COMPOUNDS AND
TITLE OF INVENTION: 79
NUMBER OF SEQUENCES: 79
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/990,571
APPLICATION NUMBER: US/08/990,571
FILING DATE: 11-DEC-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: MAKI, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.426C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 682-6031
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ. ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4223 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

[illegible]

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: February 17, 2003, 08:54:07 ; Search time 45.3031 Seconds
(without alignments)
9111.676 Million cell updates/sec

Title: US-09-605-042a-1_COPY_8000_9345

Perfect score: 1346
Sequence: 1 gacctgagatgctgctgtaaa.....gtctgcacgctgctgaccgg 1346

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5a.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5b.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5a.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/5b.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/5a.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/5b.COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	57.8	4.3	2448	2	US-08-820-170A-38
2	57.8	4.3	2448	3	US-09-055-699-38
3	57.8	4.3	2448	4	US-09-055-699-38
4	57.8	4.3	2448	4	US-09-055-699-38
5	57.8	4.3	2448	4	US-09-055-699-38
6	57.8	4.3	2448	4	US-09-055-699-38
7	57.8	4.3	2448	4	US-09-055-699-38
8	57.8	4.3	2448	4	US-09-055-699-38
9	57.8	4.3	2448	4	US-09-055-699-38
10	57.8	4.3	2448	4	US-09-055-699-38
11	57.8	4.3	2448	4	US-09-055-699-38
12	57.8	4.3	2448	4	US-09-055-699-38
13	57.8	4.3	2448	4	US-09-055-699-38
14	57.8	4.3	2448	4	US-09-055-699-38
15	57.8	4.3	2448	4	US-09-055-699-38
16	57.8	4.3	2448	4	US-09-055-699-38
17	57.8	4.3	2448	4	US-09-055-699-38
18	57.8	4.3	2448	4	US-09-055-699-38
19	57.8	4.3	2448	4	US-09-055-699-38
20	57.8	4.3	2448	4	US-09-055-699-38
21	57.8	4.3	2448	4	US-09-055-699-38
22	57.8	4.3	2448	4	US-09-055-699-38
23	57.8	4.3	2448	4	US-09-055-699-38
24	57.8	4.3	2448	4	US-09-055-699-38
25	57.8	4.3	2448	4	US-09-055-699-38
26	57.8	4.3	2448	4	US-09-055-699-38
27	57.8	4.3	2448	4	US-09-055-699-38

28	45.8	3.4	1851	2	US-08-414-657D-20	Sequence 20, Appl
29	45.8	3.4	1851	4	US-09-135-080-5	Sequence 5, Appl
30	45.8	3.4	3740	4	US-09-162-274A-6	Sequence 6, Appl
31	45.6	3.4	3925	4	US-08-793-044-1	Sequence 1, Appl
32	45.4	3.4	155	1	US-08-222-177A-19	Sequence 19, Appl
33	45.4	3.4	3854	2	US-08-720-844A-1	Sequence 1, Appl
34	45.4	3.4	3854	4	US-08-953-823A-1	Sequence 1, Appl
35	45.4	3.4	3854	4	US-09-398-239-1	Sequence 1, Appl
36	45.2	3.4	324	1	US-08-222-177A-36	Sequence 36, Appl
37	45.2	3.4	622	4	US-09-328-111-675	Sequence 675, App
38	45	3.3	2710	1	US-08-487-748A-8	Sequence 8, Appl
39	45	3.3	2710	3	US-08-480-070C-8	Sequence 8, Appl
40	45	3.3	2710	3	US-08-829-525-8	Sequence 8, Appl
41	45	3.3	2710	4	US-08-609-583A-8	Sequence 8, Appl
42	45	3.3	2710	4	US-08-937-399-8	Sequence 8, Appl
43	45	3.3	2710	4	US-09-310-367-8	Sequence 8, Appl
44	45	3.3	2710	4	US-09-032-337-8	Sequence 8, Appl
45	45	3.3	4221	1	US-07-947-120-7	Sequence 7, Appl

ALIGNMENTS

```
RESULT 1
US-08-820-170A-38
; Sequence 38, Application US/08820170A
; Patent No. 5831058
; GENERAL INFORMATION:
; APPLICANT: Tsutomu, FUJIMURA
; APPLICANT: Takeshi, MATANABE
; APPLICANT: Masato, HORIE
; APPLICANT: Toyomasa, KATAGIRI
; TITLE OF INVENTION: HUMAN GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESS: Sughrue, Mion, Zinn, Macpeak & Seas
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/820,170A
; FILING DATE:
; CLASSIFICATION: 536
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2448 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA(genomic)
; US-08-820-170A-38

Query Match 4.3%, Score 57.8; DB 2; Length 2448;
Best Local Similarity 51.1%; Pred. No. 4.6e-08;
Matches 135; Conservative 0; Mismatches 107; Indels 3; Gaps 1;

OY 918 TGTAGGACATGATGATGCTGCTACCCATGACATCACACATGCTCCACAGAGCTGT 977
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1312 TGTGAACATCATGATGATGCTGCTACAGGCGCCATTTACTGTGTGAATAATGATGT 1371
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 978 GTGAACACCCCGGCGCTGTTAAGTCTCTGTCAGATGATGTTTCTGCTGACGCCCTGAG 1037
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Db 1372 GTCACACCCCGGTTCTTTATGTCATCTGCACAACTGATACATGAAATTGAT 1431
QY 1038 CTGAGCTGCATGATGTGATGAGTCTCAGACAGGGGCTCAGTACTGATGCCCTG 1097
Db 1432 TATTCATGTACAGAAATGATGAGTGTATCACAAATCAGACAC---AATGTGATGAAAT 1488
QY 1098 GCCACCTGTCTCAACACAGAAAGGAGTCTGTCCTGTCGCCAGGGCTTACAGGG 1157
Db 1489 GCTTTATGCTTCAACACTGTTGGAGACACAATGTGTTGCAAGCGGGCTATACAGGG 1548
QY 1158 GATGG 1162
Db 1549 AATGG 1553

RESULT 2
US-09-055-699-38

; Sequence 38, Application US/09055699
; Patent No. 6005088
; GENERAL INFORMATION:
; APPLICANT: Tsutomu, FUJIWARA
; APPLICANT: Takeshi, WATANABE
; APPLICANT: Masato, HORIE
; APPLICANT: Toyomasa, KATAGIRI
; TITLE OF INVENTION: HUMAN GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/055,699
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/820,170
; FILING DATE:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2448 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA(genomic)
; US-09-055-699-38

Query Match 4.3%; Score 57.8; DB 3; Length 2448;
Best Local Similarity 55.1%; Pred. No. 4.6e-08;
Matches 135; Conservative 0; Mismatches 107; Indels 3; Gaps 1;

QY 918 TGTAGAGCATGATGATGTGCTACCCCATGAGCTCACAACACTGCTCCACAGCAGCTGT 977
Db 1312 TGTGAAGCATGATGATGTGCTGAGAGGGCGGCATTTACTGTGTAATAATACATGTGT 1371
QY 978 GTGAACACCCCGGGCTGTTAAGTCTCCTGTGAGAGTGTCTTCTGACGCCCTGAG 1037
Db 1372 GTCAACACCCCGGGTCTTTATGTGTCATCTGCAAAATGATATCATAGAAATGATGAT 1431
QY 1038 CTGAGCTGCATGATGTGATGATGCTCAGACAGGGGCTGCTAATGATGATGATGAT 1097
Db 1432 TATTCATGTACAGAAATGATGATGATGATGATGATGATGATGATGATGATGAT 1488

QY 1098 GCCACCTGTCTCAACACAGAGCGACTACTGTGCGTGTGCTCCGAGGGCTTACAGGG 1157
Db 1489 GCTTTATGCTTCAACACTGTGTGAGAGGACACACACTGTGTTCCAGCGGGCTATACAGGG 1548
QY 1158 GATGG 1162
Db 1549 AATGG 1553

RESULT 3

US-09-273-565-38
; Sequence 38, Application US/09273565A
; Patent No. 6166190
; GENERAL INFORMATION:
; APPLICANT: FUJIWARA, TSUTOMU
; APPLICANT: WATANABE, TAKESHI
; APPLICANT: HORIE, MASATO
; TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
; TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBIQUITIN-CONJUGATING ENZYME
; FILE REFERENCE: 0-53599
; CURRENT APPLICATION NUMBER: US/09/273,565A
; CURRENT FILING DATE: 1999-03-22
; EARLIER APPLICATION NUMBER: 09/055,699
; EARLIER FILING DATE: 1998-04-07
; EARLIER APPLICATION NUMBER: 08/820,170
; EARLIER FILING DATE: 1997-03-19
; EARLIER APPLICATION NUMBER: JP 63410/1996
; EARLIER FILING DATE: 1996-03-19
; EARLIER APPLICATION NUMBER: JP 69163/1997
; EARLIER FILING DATE: 1997-03-05
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 38
; LENGTH: 2448
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-273-565-38

Query Match 4.3%; Score 57.8; DB 4; Length 2448;
Best Local Similarity 55.1%; Pred. No. 4.6e-08;
Matches 135; Conservative 0; Mismatches 107; Indels 3; Gaps 1;

QY 918 TGTGAAGCATGATGATGTGCTACCCCATGAGCTCACAACACTGCTCCACAGCAGCTGT 977
Db 1312 TGTGAAGCATGATGATGTGCTGAGAGGGCGGCATTTACTGTGTAATAATACATGTGT 1371
QY 978 GTGAACACCCCGGGCTGTTAAGTCTCCTGTGAGAGTGTCTTCTGACGCCCTGAG 1037
Db 1372 GTCACACCCCGGGTCTTTATGTCATCTGCAAAATGATGATGATGATGATGATGAT 1431
QY 1038 CTGAGCTGCATGATGTGATGATGCTCAGACAGGGGCTCAGTACTGATGATGATGAT 1097
Db 1432 TATTCATGTACAGAAATGATGATGATGATGATGATGATGATGATGATGATGAT 1488
QY 1098 GCCACCTGTCTCAACACAGAAAGGAGTCTGTCCTGTCGCCAGGGCTTACAGGG 1157
Db 1489 GCTTTATGCTTCAACACTGTTGGAGACACAATGTGTTGCAAGCGGGCTATACAGGG 1548
QY 1158 GATGG 1162
Db 1549 AATGG 1553

RESULT 4

US-09-565-538-38
; Sequence 38, Application US/09565538
; Patent No. 6333404
; GENERAL INFORMATION:
; APPLICANT: FUJIWARA, TSUTOMU
; APPLICANT: WATANABE, TAKESHI
; APPLICANT: HORIE, MASATO
; TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN

us-09-605-042a-1_copy-8000_9345.rn1

```

TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBIQUITIN-CONJUGATING ENZYME
FILE REFERENCE: Q-53599
PRIORITY APPLICATION NUMBER: US/09/565,538
CURRENT FILING DATE: 2000-05-05
PRIORITY APPLICATION NUMBER: 09/273,565
PRIORITY FILING DATE: 1999-03-22
PRIORITY APPLICATION NUMBER: 09/055,699
PRIORITY FILING DATE: 1998-04-07
PRIORITY APPLICATION NUMBER: 08/820,170
PRIORITY FILING DATE: 1997-03-19
PRIORITY APPLICATION NUMBER: JP 63410/1996
PRIORITY FILING DATE: 1996-03-19
PRIORITY APPLICATION NUMBER: JP 69163/1997
PRIORITY FILING DATE: 1997-03-05
PRIORITY APPLICATION NUMBER: 2.1
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO: 38
LENGTH: 2448
TYPE: DNA
ORGANISM: Homo sapiens
US-09-565-538-38

```

RESULT 5
US-09-661-466-38
Sequence 38, Application US/09661468
Patent No. 6376189
GENERAL INFORMATION:
INVENTOR: TSUTOMU
APPLICANT: FUJIMARU, TAKESHI
APPLICANT: WATANABE, MASATO
APPLICANT: HORIE, MASATO
TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBIQUITIN-CONJUGATING ENZYME
FILE REFERENCE: Q-53599
CURRENT FILING DATE: 2000-09-13
CURRENT APPLICATION NUMBER: 09/055,659
PRIOR FILING DATE: 1998-04-07
PRIOR APPLICATION NUMBER: 08/820,170
PRIOR FILING DATE: 1997-03-19
PRIOR APPLICATION NUMBER: JP 63410/1996
PRIOR FILING DATE: 1996-03-19
PRIOR APPLICATION NUMBER: JP 69163/1997
PRIOR FILING DATE: 1997-03-05
NUMBER OF SEQ ID NOS: 95
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 38
LENGTH: 2448
TYPE DNA

US-09-661-4668-38	ORGANISM: Homo sapiens	Query Match	Best Local Similarity	Score	DB 4:	Length	2448:
Matches 135; Conservative	4.3%;	55.1%;	0;	Mismatches 107;	Indels	3;	Gaps
Dy	918	TGTGAGCATGATGATGCTGCTACCCATGACTGCACAACTGCTCCACAGAGAGCTGT	1371				
Dy	1312	TGTGAGCATGATGATGCTGCTACCCATGACTGCACAACTGCTCCACAGAGAGCTGT	1371				
Dy	978	GTCACACCCCGGCTCTTAATGCTCCTGTACAGATGCTTTTGCTGACGCTTAG	1431				
Dy	1372	GTCACACCCCGGCTCTTTTATGTGCATCTGCACAACTGATACATGATGATGAT	1097				
Dy	1038	CTGAGCTGCATGATGATGATGCTCAGACGCGGCTGCACATCTGATGCTGCTG	1488				
Dy	1432	TATTCATGTACAGACATGATGATGTATACAAATCAGCAC--AACTGTGATGAAAT	1157				
Dy	1098	GCCACCTGTGTCAACACAGAAAGGCACTCTGTGCTGTGCCGAGGCTTACAGGG	1548				
Dy	1489	GCTTTATGCTTCAACATGTTGGAGACCAACACTGTGTTTGCAACCGGCTATACAGG	1548				
Dy	1158	GATGG	1162				
Dy	1549	AATGG	1553				

2020-170A-39
ence 39, Application US/08820170A
nt No. 5831058
ERAL INFORMATION: FUJIWARA
PLICANT: TSUCOMU, WATANABE
PLICANT: TAKESEHI, HORIE
PLICANT: MASAO, KATAGIRI
PLICANT: TOYOMASA, HUMAN GENE
TITL E OF INVENTION: 42
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS: 42
ADDRESSEE: Sugitru, Mion, Zinn, Macpeak & Seas
STREET: 2100 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: United States
ZIP: 20037-3202
COMPUTER READABLE FORM:
MED IUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.300
CURRENT APPLICATION NUMBER: US/08/820,170A
APPLICATION DATE:
CLASSIFICATION: 536
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 293-7060
TELEFAX: 6491103
TELEX: 6491103
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 3198 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA(genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE: Human fetal brain cDNA library
LIBRARY: GEN-093E05
CLONE:
NAME/KEY: CDS

LOCATION: 97...2544
000-170A-39

LOCATION:	DB 2:	Length	3198:
S-08-820-170A-39			
Query Match	4.38;	Score 57.8;	
Best local Similarity	55.18;	Pred. No. 5,3e-08;	
Matches 135;	Conservative	Mismatches 107;	Indels 3; Gaps 1
<p> ...ATGAGAGTGTCTACCCAGACGTCAACATCTGCTCCACAGCAGCGTGT ...ATGAGAGTGTCTACCCAGACGTCAACATCTGCTCCACAGCAGCGTGT 1467 </p>			

[illegible]

RESULT 7
US-09-055-699-39
US/09055699

Patent NO. 6005086
GENERAL INFORMATION: FUJIMARA
APPLICANT: Tsumoto, MATANABE
APPLICANT: TAKESHI, HORIE
APPLICANT: Masato, KATAGIRI
APPLICANT: Toyomasa, HUMAN
TITLE OF INVENTION: 42
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS: Mito, Yama, Macpeak & Seas
ADDRESSEE: Sughrue, Mito, Yama, Macpeak & Seas
STREET: 2100 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: United States
ZIP: 20037-3702
COMPUTER READABLE FORM:
MEDIA TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA: 05/09/05, 699
APPLICATION NUMBER: 05/09/05, 699

CLASSIFICATION DATA: 08/820, 170
PRIOR APPLICATION NUMBER: 08/820, 170
APPLICATION NUMBER: 08/820, 170
FILING DATE: 08/820, 170
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 293-7860
TELEFAX: (202) 293-7860
TELEX: 6491103
ID NO: 39:
INFORMATION FOR SEQ. ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 3198 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA(genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE: Human fetal brain cDNA library
LIBRARY:

CLONE: GEN-093E05
FEATURE: CDS
NAME/KEY: 97...2544
LOCATION: 39
US-09-055-699-39

LOCAL	LOCAT	Score	DB:3	Length	3198
S-09-055-699-39		4.3%	Pred. No. 5.3e-08	Indels	3
Query Match		55.18		Gaps	1
Best Local Similarity		0	Mismatches	107	
Matches	135	Conservative			
<p> ...GCGACCCACGATGACGTCACACACTGCTCCACAGCAGCTGT ...GCGACCCACGATGACGTCACACACTGCTCCACAGCAGCTGT </p>					

[illegible]

RESULT 8
US-09-273-565-39
Application US/09273565A

```

1  GENERAL INFORMATION:  TSTOMU
2  APPLICANT:  FUJIMARU, TAKESHI
3  APPLICANT:  MATSUNABE, TAKESHI
4  APPLICANT:  HORIE, MASARU
5  TITLE OF INVENTION:  AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
6  TITLE OF INVENTION:  SKELETAL MUSCLE-SPECIFIC UBIQUITIN-CONJUGATING ENZYME
7  FILE REFERENCE:  0-53599
8  CURRENT FILING DATE:  1999-03-22
9  CURRENT FILING DATE:  1999-03-22
10 CURRENT FILING DATE:  1999-03-22
11 EARLIER FILING DATE:  1998-04-07
12 EARLIER FILING DATE:  1998-04-07
13 EARLIER FILING DATE:  1998-04-07
14 EARLIER FILING DATE:  1997-03-19
15 EARLIER FILING DATE:  1997-03-19
16 EARLIER FILING DATE:  1996-03-19
17 EARLIER FILING DATE:  1996-03-19
18 EARLIER FILING DATE:  1997-03-05
19 EARLIER FILING DATE:  1997-03-05
20 EARLIER FILING DATE:  1997-03-05
21 NUMBER OF SEQ ID NOS:  95
22 SOFTWARE:  Patentin Ver. 2.1
23 SEQ ID NO 3198
24 LENGTH:  3198
25 TYPE:  DNA
26 ORGANISM:  Homo sapiens
27 FEATURE:
28 NAME/KEY:  CDS
29 LOCATION:  (97)..(2544)
30 US-09-273-565-39
31 4.3%; Score 57.8; DB 4; Length 3198;
32 3; Gaps
33 1

```

[illegible]

ns-09-605-042a-1_copy-8000_9345.rni

RESULT 9
US-09-565-538-39
Application US/09565538

Patent No. 333404
 GENERAL INFORMATION: TSUTOMU
 APPLICANT: FUJIWARA, TAKESHI
 APPLICANT: MATSUNABE, TAKESHI
 APPLICANT: HORIE, MASATO
 TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
 SKELETAL MUSCLE-SPECIFIC UBIOUITIN-CONUGATING ENZYME
 TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBIOUITIN-CONUGATING ENZYME
 FILE REFERENCE: Q-53599
 CURRENT APPLICATION NUMBER: US/09/565,538
 CURRENT FILING DATE: 2000-05-05
 CURRENT FILING DATE: 09/27/3,565
 PRIOR APPLICATION NUMBER: 09/055,699
 PRIOR FILING DATE: 1999-03-22
 PRIOR APPLICATION NUMBER: 08/820,170
 PRIOR FILING DATE: 1998-04-07
 PRIOR APPLICATION NUMBER: JP 63410/1996
 PRIOR FILING DATE: 1997-03-19
 PRIOR APPLICATION NUMBER: JP 69163/1997
 PRIOR FILING DATE: 1996-03-19
 PRIOR APPLICATION NUMBER: JP 69163/1997
 PRIOR FILING DATE: 1997-03-05
 NUMBER OF SEQ ID NOS: 95
 SOFTWARE: Patentln Ver. 2.1
 SEQ ID NO 39
 LENGTH: 3198
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (97)..(2544)
 US-09-565-538-39

[illegible]

```

RESULT 10
US-09-661-468-39 Application US/09661468
Sequence 39, Patent No. 6376189
GENERAL INFORMATION:
APPLICANT: FUJIWARA, TSUTOMU
APPLICANT: MATSUNABE, TAKESHI
APPLICANT: HORIE, MASAO
APPLICANT: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBIQUITIN-CONJUGATING ENZYME
TITLE OF INVENTION: 0-53599
FILE REFERENCE: US/09/661,468
CURRENT APPLICATION NUMBER: 2000-09-13
CURRENT FILING DATE: 09/05/5, 699
PRIORITY APPLICATION NUMBER: 08/820,170
PRIORITY FILING DATE: 1997-03-19
PRIORITY APPLICATION NUMBER: 63410/1996
PRIORITY FILING DATE: 1996-03-19
PRIORITY APPLICATION NUMBER: JP 69163/1997
PRIORITY FILING DATE: 1997-03-05
NUMBER OF SEQ ID NOS: 95
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 39
LENGTH: 3198
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (97)..(2544)
US-09-661-468-39

```

Query Match	4.3%	Pred. No. 5.3e+08	Indels	3	Gaps
Best Local Similarity	55.1%	0	Mismatches	107	
Matches 155	Conservative				
QY	918	TGTGAGACATGTGATAGTGTGCTACCCCATGGATGCACACACTGCTCCACAGCAGCTGT	1467		
Db	1408	TGTAAAGCATCATGATGATGTGCTGAAGGGCCCATTTACTTGCTGTGACCCCTGAG	1037		
QY	978	GTGAACACCCCGGGCTCGTTAAAGTCCCTCTCAGAGATGTTTGTGCTGACCCCTGAG	1527		
Db	1468	GTCAACACCCCGGGCTCTTTATGTGCAATGTGCAAAACTGGATATCATCAGAAATGATAT	1097		
QY	1038	CTGAGCTGCATGATGTGATGATGATGCTCAGACGAGGGCTCAGTACTGTGATGCCCTG	1584		
Db	1528	TATTCATGTATACAGACATGATGATGTATACAAATACAGAC--ACTGTGATGAAAT	1157		
QY	1098	GCCACCTGTGTCAACACAGAGGCGACTCTGTGCGTGTGTCCGAGGGCTTACAGGG	1644		
Db	1585	GCTTATGCTTCAACACTGTTGGAGGACACAACTGTTGTTCAGAGCGGGGCTATACAGG	1644		
QY	1158	GATGG 1162			
Db	1645	ATGG 1649			
RESULT 11					
US-08-637-823B-26					
; Sequence 26, Application US/08637823B					
; Patent No. 6184031					
; GENERAL INFORMATION:					
; APPLICANT: Gros, Philippe					
; APPLICANT: Skanene, Emil					
; TITLE OF INVENTION: DNA SEQUENCES THAT ENCODE A NATURAL					
; TITLE OF INVENTION: RESISTANCE TO INFECTION WITH INTRACELLULAR PARASITES					
; NUMBER OF SEQUENCES: 32					
; CORRESPONDENCE ADDRESS: KLAUBER & JACKSON					
; ADDRESS: 411 Hackensack Ave					
; STREET: 411 Hackensack Ave					
; CITY: Hackensack					
; STATE: New Jersey					

[illegible]

```

RESULT 14
US-09-422-936-60/C, Application US/09422936
Sequence 60, Application 6465213
GENERAL INFORMATION:
APPLICANT: Ekstrand, Jonas
PATENT No. 6465213
TITLE OF INVENTION:
FILE REFERENCE: 06275-165002
CURRENT APPLICATION NUMBER: US/09/422,936
CURRENT FILING DATE: 1999-10-22
PRIOR APPLICATION NUMBER: US/09/242,608
PRIOR FILING DATE: 1999-02-19
PRIOR APPLICATION NUMBER: SWEDEN 9703914-2
PRIOR FILING DATE: 1998-10-27
PRIOR APPLICATION NUMBER: SWEDEN 9800864-2
PRIOR FILING DATE: 1997-10-27
PRIOR APPLICATION NUMBER: SWEDEN 9800864-2
PRIOR FILING DATE: 1998-03-16
PRIOR APPLICATION NUMBER: SWEDEN 9802575-2
PRIOR FILING DATE: 1998-07-17
NUMBER OF SEQ ID NOS: 85
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO: 60
LENGTH: 15652
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc-feature
LOCATION: (1)..(15652)
OTHER INFORMATION: n = A, T, G, or C
NAME/KEY: misc-feature
LOCATION: (1)..(15652)
OTHER INFORMATION: f = G or A; y = T/U or C; m = A or C;
OTHER INFORMATION: k = G or T/U; s = G or C; w = A or T/U;
OTHER INFORMATION: b = G, C, or T/U; d = A, G, or C
OTHER INFORMATION: h = A, C, or T/U; v = A, G, or C
NAME/KEY: exon
LOCATION: (3419)..(3444)
NAME/KEY: intron
LOCATION: (3445)..(3908)
NAME/KEY: exon
LOCATION: (3909)..(3993)
NAME/KEY: intron
LOCATION: (3994)..(4694)
NAME/KEY: exon
LOCATION: (4695)..(4898)
NAME/KEY: intron
LOCATION: (4899)..(5652)
NAME/KEY: exon
LOCATION: (5653)..(5838)
NAME/KEY: intron
LOCATION: (5839)..(7184)
NAME/KEY: exon
LOCATION: (7185)..(7205)
NAME/KEY: intron
LOCATION: (7206)..(8310)
NAME/KEY: exon
LOCATION: (8311)..(8806)
NAME/KEY: intron
LOCATION: (8807)..(12271)
NAME/KEY: exon
LOCATION: (12272)..(12406)
NAME/KEY: intron
LOCATION: (12407)..(12820)
NAME/KEY: exon

```

```
LOCATION: (12821)...(12991)
NAME/KEY: intron
LOCATION: (12992)...(14089)
NAME/KEY: exon
LOCATION: (14090)...(14191)
NAME/KEY: intron
LOCATION: (14192)...(14477)
NAME/KEY: exon
LOCATION: (14478)...(14543)
NAME/KEY: intron
LOCATION: (14544)...(15002)
NAME/KEY: exon
LOCATION: (15003)...(15194)
NAME/KEY: intron
LOCATION: (15195)...(15652)
NAME/KEY: intron
LOCATION: (15653)...(15652)
```

[illegible]

RESULT 15
 US-08-675-773B-3/5 Application US/08675773B
 Sequence 3, Application
 Patent No. 6166288
 GENERAL INFORMATION:
 APPLICANT: DIAMOND, LISA E
 APPLICANT: LOGAN, JOHN S
 APPLICANT: BYRNE, GUERARD W
 APPLICANT: SHARRM, AJAY
 TITLE OF INVENTION: METHOD OF PRODUCING TRANSGENIC ANIMALS
 TITLE OF INVENTION: FOR XENOTRANSPLANTATION.
 NUMBER OF SEQUENCES: 42
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: BROMDY AND NETMARK, P.L.L.C.
 STREET: 419 SEVENTH STREET, N.W., SUITE 300
 CITY: WASHINGTON
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20004
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION NUMBER: US/08/675,773B
 FILING DATE: 03-JUL-1996
 CLASSIFICATION: 800
 ATTORNEY/AGENT INFORMATION:
 NAME: COOPER, IVER P. 005
 REGISTRATION NUMBER: 28, 005
 REFERENCE/DOCKET NUMBER: DIAMOND=1A
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-628-5197
 TELEFAX: 202-737-3528
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3791 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 17, 2003, 08:33:57 : Search time 1720.48 Seconds
(Without alignments)
12670.391 Million cell updates/sec

Title: US-09-605-042a-1_COPY_8000_9345
Perfect score: 1346
Sequence: 1 gaccctgcagctgcgtgta.....gtctgcacgctgtaccgg 1346

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 809774376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hrc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rnd:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	526.2	39.1	656	12	BF783842
2	526.2	39.1	703	13	BI100241
3	526.2	39.1	787	12	BF789447
4	526.2	39.1	1087	12	BF785672
5	524.4	39.0	906	12	BF783809
6	515.2	38.3	801	13	BG969253
					BG969253 602836721

7	515.2	38.3	808	13	BG972563
8	515.2	38.3	832	13	BG969349
9	515.2	38.3	964	12	BF789128
10	513.6	38.2	564	9	AI527836
11	513.2	38.1	686	10	AM211279
12	509.4	37.8	608	9	AI527884
13	505.6	37.6	713	9	AA386595
14	504.2	37.5	940	12	BF785008
15	504	37.4	843	12	BF786633
16	501.2	37.2	1054	13	BF781466
17	500.6	37.2	622	13	BG970639
18	500.2	37.2	585	9	AI747914
19	499.6	37.1	593	9	AI747396
20	497.8	37.0	969	12	BF687119
21	497.4	37.0	888	12	BF783802
22	495.2	36.8	672	12	BF789723
23	494.2	36.7	769	12	BF787946
24	493.6	36.7	790	13	BI100943
25	492.6	36.6	914	12	BF788589
26	485.6	36.1	683	13	BI101569
27	485.2	36.0	754	13	BI103612
28	484.6	36.0	754	9	AI987598
29	483.6	35.9	640	12	BF786743
30	479.6	35.6	1078	12	BF780526
31	477	35.4	1042	12	BF786518
32	476.2	35.4	861	12	BF786084
33	474	35.2	749	12	BF788214
34	471.6	35.0	646	12	BF789452
35	471.4	35.0	580	10	AM106846
36	470.4	34.9	852	12	BF786238
37	470.2	34.9	584	12	BF53234
38	469	34.8	707	13	BI100595
39	467.8	34.8	584	9	AI663313
40	465.8	34.6	909	12	BF786646
41	462	34.3	682	9	AI746802
42	462	34.3	1118	12	BF687200
43	461.8	34.3	1090	12	BF781608
44	455.6	33.8	554	9	AI746618
45	455	33.8	871	12	BF784752

ALIGNMENTS

RESULT 1
LOCUS BF783842 656 bp mRNA linear EST 12-JAN-2001
DEFINITION 602109765F1 NCL CGAP_Kid14 Mus musculus CDNA clone IMAGE:4237744
5', mRNA sequence.
ACCESSION BF783842
VERSION BF783842.1 GI:12088878
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL 1 (bases 1 to 656)
NII-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
CDNA Library Preparation: Jeffrey E. Green, M.D.
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9848 row: b column: 17
High quality sequence start: 2
High quality sequence stop: 654.
Location/Qualifiers

FEATURES

JOURNAL
COMMENT
unpublished
Contact: Robert Strausz, n.h.gov
Email cgabs-remail n.h.gov

10


```
|||||
Db 471 TGTCTCCCAAGGCTCCGTGTAGCCAGACTGAGCTGCTTGTCCCAAGGCGCCGATGGAAG 530
QY 1236 CTGCTGTGTCAAGACCCCTGCATATATAGAGCCCTGACTGAGTACTGGCGACACA 1295
Db 531 CTGCTGTGTCAAGACCCCTGCATATATAGAGCCCTGACTGAGTACTGGCGACACA 590
QY 1296 GAGTATGCTGTGGCTACTCTCTGTGACGGGCTGTGCACGGCTGTGACCGG 1346
Db 591 GAGTATGCTGTGGCTACTCTCTGTGACGGGCTGTGCACGGCTGTGACCGG 641

RESULT 5
BF783809 906 bp mRNA linear EST 12-JAN-2001
LOCUS 602109729p1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:4237926
DEFINITION 5', mRNA sequence.
ACCESSION BF783809
VERSION BF783809.1 GI:12088845
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 906)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.llnl.gov
Plate: LHAM9848 row: j column: 07
High quality sequence stop: 670.
Location/Qualifiers
1. 906
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_image="IMAGE:4237926"
/clone_lib="NCI_CGAP_Kid14"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: Kidney; Vector: pCMV-SPORT6; Site: 1; NotI;
Site: 2; SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library. |"
BASE COUNT 170 a 244 c 290 g 201 t 1 others
ORIGIN
Query Match 39.0%; Score 524.4; DB 12; Length 906;
Best Local Similarity 99.8%; Pred. No. 2.1e-145;
Matches 525; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 821 ACGGCTCTGTGATCCCAACACCCCACTGCACGGTGGATGGTGTGCACAACTG 880
Db 1 ACGGCTCTGTGATCCCAACACCCCACTGCACGGTGGATGGTGTGCACAACTG 60
QY 881 CTCTGACAGACCGGCTTCACTGTGTATGGGCTGTGTGTGAGACATGATGATGTC 940
Db 61 CTCTGACAGACCGGCTTCACTGTGTATGGGCTGTGTGTGAGACATGATGATGTC 120
QY 941 TACCCCAAGGACTCACAACTGTCTCCAAACAGACACTGTGTGAACCCCGGCTGTTAA 1000
Db 121 TACCCCAAGGACTCACAACTGTCTCCAAACAGACACTGTGTGAACCCCGGCTGTTAA 180
QY 1001 GTGCTCTGTGTGAGATGGTTTGTGTGTGAGAGCTGAGCTGAGCTGATGTGATGA 1060
Db 181 GTGCTCTGTGTGAGATGGTTTGTGTGTGAGAGCTGAGCTGATGTGATGA 240

FEATURES
source
1. 906
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_image="IMAGE:4237926"
/clone_lib="NCI_CGAP_Kid14"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: Kidney; Vector: pCMV-SPORT6; Site: 1; NotI;
Site: 2; SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library. |"
Location/Qualifiers
1. 906
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_image="IMAGE:4237926"
/clone_lib="NCI_CGAP_Kid14"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: Kidney; Vector: pCMV-SPORT6; Site: 1; NotI;
Site: 2; SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library. |"
BASE COUNT 170 a 244 c 290 g 201 t 1 others
ORIGIN
```

```
QY 1061 GTGCTCAGACGAGGGGCTCAGTACTGTATGCTTGGCCCACTGTGTCAACAGAAAG 1120
Db 241 GTGCTCAGACGAGGGGCTCAGTACTGTATGCTTGGCCCACTGTGTCAACAGAAAG 300
QY 1121 CGACTACTTGGCGGTGTCCCGAGGGCTTTACAGGGATGTTGTGACTGTGATGCTC 1180
Db 301 CGACTACTTGGCGGTGTCCCGAGGGCTTTACAGGGATGTTGTGACTGTGATGCTC 360
QY 1181 CCCAGGCTCTGTGAGACCCAGGACTGCTTGTGCCCCAGGGCCGATGGAAGCTGCT 1240
Db 361 CCCAGGCTCTGTGAGACCCAGGACTGCTTGTGCCCCAGGGCCGATGGAAGCTGCT 420
QY 1241 GTGTCAAGACCCCTGCATATATAGAGCCCTGACTGACTGTGGCGACAGAGTA 1300
Db 421 GTGTCAAGACCCCTGCATATATAGAGCCCTGACTGACTGTGGCGACAGAGTA 480
QY 1301 TGGTGTGGGCTACTCTGTGTGACGGGGTCTGCACGGCTGTGACCGG 1346
Db 481 TGGTGTGGGCTACTCTGTGTGACGGGGTCTGCACGGCTGTGACCGG 526

RESULT 6
BG969253 801 bp mRNA linear EST 12-JUN-2001
LOCUS 602836721F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:4971340
DEFINITION 5', mRNA sequence.
ACCESSION BG969253
VERSION BG969253.1 GI:14356890
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 801)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.llnl.gov
Plate: LHAM0956 row: i column: 05
High quality sequence stop: 768.
Location/Qualifiers
1. 801
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_image="IMAGE:4971340"
/clone_lib="NCI_CGAP_Kid14"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: Kidney; Vector: pCMV-SPORT6; Site: 1; NotI;
Site: 2; SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library. |"
BASE COUNT 165 a 209 c 246 g 181 t
ORIGIN
Query Match 38.3%; Score 515.2; DB 13; Length 801;
Best Local Similarity 99.2%; Pred. No. 1.1e-142;
Matches 528; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 816 GCAAGAGCGGTGTGTGATGCCACAACACCGCACTGCACGGTGTGATGGTGTACA 875
Db 174 GCAAGAGCGGTGTGTGATGCCACAACACCGCACTGCACGGTGTGATGGTGTACA 233
QY 876 ACGTCTCTCTGCAAGACCGGCTTCACTGCTGTATGGGCTGTGTGAGACATGATGAG 935
```


US-09-60-57

BASE COUNT		
ORIGIN	100 d	DB 13; Length 832;
	score 515.2;	

Average. Note: this is a 1 others
Technologies. 196 t
244 c 308 g
215 a

1

0

2

2

45

11

ORGANISM Mus musculus
Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Sciuromorphi; Muridae; Murinae; Mus
Rodentia;

TITLE The Wasnu-NCI Mouse
JOURNAL Unpublished (1999)
COMMENT Contact: Maira M/Wasnu-NCI Mouse EST Project 1999
Washington University School of Medicine
Box 8501, St. Louis, MO 63108, USA

Fax: 314 286 1010
Email: mouse@station.wustl.edu
This clone is available royalty-free through LLN; contact the IMAGE Consortium (info@image.lln.gov) for further information.
MGI:977690
Seq. primer: custom primer used
High quality sequence stop: 318.
Information/Qualifiers

	SOURCE	1.	564	/organism="Mus musculus"
				/strain="C57BL"
				/db_xref="taxon:10090"
				/clone="IMAGE:1921398"
				/clone_1lb="Sugano mouse kidney mklta"
				/sex="female"
				/dev_stage="adult"
				/lab_host="DH10B"
				/note="Organ: kidney; Vector: pME18-FL3; Site:1: DraIII /CACCTGTG); Site:2: DraIII (CACCATGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGGCGCTTTTCTTTTTTTTTT]; double-stranded cDNA was ligated into distinct draIII sites of the pME18-FL3 and cloned site CACCTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end Primer CTTCCTCTCTTAAGACTGG and 3' end primer CGAAGCTGCAGCTTGACCA."
BASE COUNT		111 a	150 c	180 g 123 t
ORIGIN				
Query Match	38.2%	Score 513.6;	DB: 9;	Length 564;
Best Local Similarity	99.1%;	Pred. No. 3e-142;	4;	Gaps 1.
Matches 527;	Conservative	0;	Mismatches	
OY	816 GCAGACGCGTGTCTGAATGCCAACCAACACAGCCACCTGCACGCGTAGAGTGGTGTACA	875		
Db	11			
	12 GCGAGACGCGTGTCTGAATGCCAACCAACAGCCACCTGCACGCGTAGAGTGGTGTACA	71		
OY	876 AGTGCTCTCCGACGACCAGCGCTTCACTACGTGGATGGGGTGGTGTGTAGSACATGGATTAG	935		
Db	72 ACCTGCTCTCCGACGACCAGCGCTTCACTACGTGGATGGGGTGGTGTGTAGSACATGGATTAG	131		
OY	936 TGTGCTACCCCATGAGCTACAACTGCTCAAAGAGAGAGCTGTGAACA -CCCCGGCTC	994		
Db	132 TGTGCTACCCCATGAGCTACAACTGCTCAAAGAGAGAGCTGTGTGAACACCCC GGCTC	191		
OY	995 GTTTAACTGCTCTGTCTCAGAGATGGTTTTCGTGACGCCCTGAGCTGGAAGTCACTGATGT	1054		
Db	192 GTTTAAAGCTCTGTCTCAGAGATGGTTTTCGTGACGCCCTGAGCTGGAAGTCACTGATGT	251		
OY	1055 GGATAGTGTCTCAAGACAGAGGGCTCAGTAACTCATGCTCTGGCCCTGACCTGTGTCAACAC	311		
Db	252 GGATAGTGTCTCAAGACAGAGGGCTCAGTAACTCATGCTCTGGCCCTGACCTGTGTCAACAC	1174		
OY	1115 AAGAAGCGACTCTGTGCTGTGTGCCGAGAGGCTTTACAGAGGATGTTGGTACTGTGCA	1174		

RESULT 11		686 bp	mRNA	linear	EST 03-DEC-1993
AW211279					
LOCUS	AW211279				
DEFINITION	u129h09.y1 Sugano mouse kidney mXia mus musculus CDNA clone				
	IMAGE:2099777	5', similar to gb:MI5881 UROMODULIN PRECURSOR (HUMAN); mRNA			
	gb:133406	Mus musculus uromodulin mRNA, complete cds (MOUSE);			
		sequence.			
ACCESSION	AW211279				
VERSION	AW211279.1	GI:6517227			
KEYWORDS					
SOURCE	house mouse,				
ORGANISM	mus musculus,				
	EST.				
	Eukaryota; Metazoa; Chordata; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
	Mammalia; Eutheria;				
	1 (bases 1 to 686)				
REFERENCE	Marras,M., Hillier,L., Kucaba,T., Martin,J., Beck C., Wylie,T., Person				
AUTHORS	Marras,M., Hillier,L., Kucaba,T., Martin,J., Beck C., Wylie,T., Person				
	Underwood,K., Stepien,M., Pfeisinger,D., Harvey,N., McCann,R.,				

TITLE
 JOURNAL
 COMMENT
 The unpublished (1999)
 Other ESTs: ul29h09.x1
 Contact: Maria M/Mashu-NCI Mouse EST Project 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 IMAGE:996709
 MGI:996709
 Seq primer: custom primer used
 High quality sequence stop: 377.
 Location/Qualifiers
 1..686
 1. :organism="Mus musculus"
 /strain="C57BL"
 /db_xref="taxon:10090"
 /clone="IMAGE:2099777"
 /clone_11b="Sugano mouse kidney mk1a"
 /sex="female"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
 (CACGCTGG); Site_2: DraIII (CACCAGTGG); 1st strand cDNA
 (CACGCTGG); double-stranded cDNA was
 was primed with an oligo(dT) primer
 [AGTGGCCCTTTTTTTTTTTTTT]; (TGTGGGCCACCTGG), digested
 ligated to a DraIII adaptor (the pME18S-FL3
 and cloned into distinct DraIII sites of the pME18S-FL3
 vector (5' site CACGCTGG, 3' site CACCAGTGG). XhoI should
 be used to isolate the cDNA insert. Size selection was
 performed to exclude fragments <1.5kb. Library
 constructed by Dr. Sumio Sugano (University of Tokyo
 Institute of Medical Science). Custom primers for
 sequencing: 5' end primer CTGCTGCTAAAGCTCG and 3' end
 primer CGACCTGCAGCTGACGACA."

MGI: 977753
Seq primer: custom primer used
High quality sequence stop: 329.
Location/Qualifiers

FEATURES
source

```

BASE COUNT
ORIGIN
119 a 163 c 194 g 131 t 1 others
27 aa: Score 509.4; DB: 9; Length 608;
1

```

RESULT 12
 LOCUS A1527894
 DEFINITION
 A1527894
 u330d11.y1 Sugano mouse kidney m2ia Mus musculus utromodilin mRNA.
 IMAGE:1921461 5' similar to gb:13406 Mus musculus utromodilin mRNA.
 complete cds (MOUSE);, mRNA sequence.
 A1527894
 A1527894.1 GI:4442029
 house mouse.
 EST.
 Mus musculus
 Eukaryota: Metazoa: Chordata: Cranialia: Vertebrata: Euteleostomi;
 Mammalia: Eutheria: Rodentia; Sciurognathi: Muridae; Murinae: Mus
 1 (bases 1 to 608) Kuwabata,T., Martin,J., Beck,C., Wylie,T., Perso
 Marra,M., Hillier,L., Kucaba,T., Allen,N., Bowers,Y., Schurk,R., Ritten
 Underwood,K., Steptoe,M., Pfeifer,B., Hadden,N., Schurk,R., Ritten
 'B., Smaller,T., Gibbons,M., Page,D., Harvey,N., McCann,R.,
 E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
 Waterston,R. and Wilson,R.
 The WashU-NCI Mouse EST Project 1999
 Unpublished (1999)
 Contact: Marita M/WashU-NCI Mouse EST Project 1999
 Washington University Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1810
 Fax: 314 286 1810
 Email: mouseest@tason.wustl.edu
 This clone is available royalty-free through LINT.
 IMAGE Consortium (info@image.llnl.gov) for further information.

DEFINITION	uc81910.y1 Sugano mouse kidney mRna Mus musculus cDNA clone IMAGE:1432098 5' similar to gb:M15881 UROMODULIN PRECURSOR (HUMAN); gb:D3406 Mus musculus uromodulin mRNA, complete cds (MOUSE);, mRNA sequence.					
ACCESSION	AA986595					
VERSION	AA986595.1 GI:3168249					
KEYWORDS	EST.					
SOURCE	house mouse.					
ORGANISM	Mus musculus.					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus. 1 (bases 1 to 713)					
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubouque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,R., Tan,F., Underwood,R., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.					
TITLE	The WashU-HMI Mouse EST Project					
JOURNAL	Unpublished (1996)					
COMMENT	Contact: Marra M/Mouse EST Project WashU-HMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@wustl.edu This clone is available royalty-free through LML; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:916166 Seq primer: primer name ambiguous High quality sequence stop: 349. Location/Qualifiers 1..713 /organism="Mus musculus" /strain="C57BL" /db_xref="taxon:10090" /clone="IMAGE:1432098" /clone_1fb="Sugano mouse kidney mRna" /sex="female" /dev_stage="adult" /lab_host="DH10B" /note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII (CACTGTG); Site_2: DraII (CACCATGTG); 1st strand CDNA was primed with an oligo(dT) primer [ATGGGCCCTTTTTTTTTTTTTTTT]; double-stranded CDNA was ligated to a DraIII adaptor [TGTGGCCCTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Sciences). Custom primers for sequencing: 5' end primer CTTCCTCTATAAGCTGG and 3' end primer CGACCTGCAGCTCGAGACA." BASE COUNT 141 a 189 c 230 g 152 t 1 others ORIGIN					
Query Match	37.6%	Score 505.6;	DB 9:	Length 713;		
Best Local Similarity	99.0%	Pred. No. 8.2e+140;				
Matches 519;	Conservative	0;	Mismatches	4;	Indels	1; Gaps 1;
Df	816	GCAGAGCGGATTCTGAATGCACACACAGCGCACCTGCAGCGTGATGGTGTGTGTACA	875			
Df	12	GCGAGACGGGTCTTGATGTGCACACACACGCGCACCTGCAGCGTGATGGTGTGTGTACA	71			
Df	876	ACGTGCTCCTGCAGACCGGGCTTCACGTGGTGTATGGGCTGTGTGTAGGACATGATGAG	935			
Df	72	ACGTGCTCTCCGACAGACCGGCTTCACGTGGTGTATGGGCTGTGTGTAGGACATGATGAG	131			
Df	936	TGTGTACCCCAAGATCACAATCTGCTCCACACAGACAGTGTGTGTAAACACCCGGGCTCG	995			
Df	132	TGTGTACCCCAAGATCACAATCTGCTCCACACAGACAGTGTGTGTAAACACCCGGGCTCG	191			

QY	816	GCAGAGCGGTCTTCAATGCACCAACACGACCCACACTGCAGCGTGGATGGTGGTCACCA	875
QY	996	TTTAAAGTCCTCCTCAGAGATGGTTTGGTCTGAGCCCTGAGCGTGCAGCTGCACATGATGTG	1055
Db	192	TTTAAAGTCTCCTCCTCAGAGATGGTTTGGTCTGAGCCCTGAGCGTGCAGCTGCACATGATGTG	251
QY	1056	GATGAGTGTCTCAGAGCAGAGGGGCTCAGTAATCTGCATGCCCTTGCCACACTGTGTCAACACA	1115
Db	252	GATGAGTGTCTCAGAGCAGAGGGGCTCAGTAATCTGCATGCCCTTGCCACACTGTGTCAACACA	311
QY	1116	GAAGCGACTACTTGTGCGTGTGTCCCGAGAGGCTTTACAGGGGATGGTGGTACTGTGAG	1175
Db	312	GAAGCGACTACTTGTGCGTGTGTCCCGAGAGGCTTTACAGGGGATGGTGGTACTGTGAG	371
QY	1176	TGCTCCCGAGGCTCCTGTGAGCCAGAGACTGAGACTGCTTGTCCCGAGAGG-CCCGATGTGAAA	1234
Db	372	TGCTCCCGAGGCTCCTGTGAGCCAGAGACTGAGACTGCTTGTCCCGAGAGG-CCCGATGTGAAA	431
QY	1235	GCATGCTGTCAAGACCCCTGCAATACATATATAGACCTGTGACTGAGTATGGCGACACAC	1294
Db	432	GCATGCTGTGTCAAGACCCCTGCAATACATATAGACACCTGTGACTGAGTATGGCGACACAC	491
QY	1295	AGAGTATGTGTGGGCTACTCCTGTGTGACGCGGGGTGTGACGCGCT	1338
Db	492	AGAGTATGTGTGGGCTACTCCTGTGTGACGCGGGGTGTGACGCGCT	535
RESULT 14			
LOCUS	BF785008	602110958F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:4338942	940 bp mRNA linear EST 12-JAN-2001
DEFINITION		5', mRNA sequence.	
ACCESSION	BF785008		
VERSION	BF785008.1	GI:12090031	
KEYWORDS		EST.	
SOURCE		house mouse.	
ORGANISM		Mus musculus	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
TITLE		1 (Bases 1 to 940)	
JOURNAL		NIH-MGC http://mgc.nci.nih.gov/ .	
COMMENT		National Institutes of Health, Mammalian Gene Collection (MGC)	
		Unpublished (1999)	
		Contact: Robert Strausberg, Ph.D.	
		Email: cgapbs-remail.nih.gov	
		Tissue Procurement: Jeffrey E. Green, M.D.	
		CDNA Library Preparation: Life Technologies, Inc.	
		CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)	
		DNA Sequencing by: Incyte Genomics, Inc.	
		Clone distribution: MGC clone distribution information can be	
		found through the I.M.A.G.E. Consortium/LNLN at:	
		http://Image.lnl.gov	
		Plate: LHAM9851 row: d column: 15	
		High quality sequence stop: 675.	
FEATURES		Location/Qualifiers	
source		1..940	
		/organism="Mus musculus"	
		/strain="FVB/N"	
		/db_xref="taxon:10090"	
		/clone_image="IMAGE:4338942"	
		/clone_lib="NCI_CGAP_Kid14"	
		/lab_host="DH10B (TI phage-resistant)"	
		/note="Organ: kidney; vector: pCMV-Sport6; Site_1: NotI;	
		Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.	
		Average insert size 1.75 kb. Constructed by Life	
		Technologies. Note: this is a NCI_CGAP Library. 1"	
BASE COUNT	201 a	249 c	314 g 176 t
ORIGIN			
Query Match	37.5%	Score 504.2;	DB 12; Length 940;
Best Local Similarity	99.1%	Pred. No. 2.4e-139;	
Matches 528; Conservative	0;	Mismatches 3;	Indels 2; Gaps 2;

Wed Feb 19 14:09:19 2003

us-09-605-042a-1_copy_8000_9345.rst

Db 122 GCGAGACGGTGTCTGATATCCACAAACAGCCACCTGCACGGTGTGATGTTGTCACA 181
 876 AC-GTGTCTCCGCGACAGCCGGCTTCACTGTGTATGGGGCTGTGTGAGACATGATGTA 934
 187 ACGGTGCTCCGCGACAGCCGGCTTCACTGTGTATGGGGCTGTGTGAGACATGATGTA 241
 935 GTGTGCTACCCATGATGATCACTGCTCCAAACAGCAGCTGTGTGACACCCGGGCTC 994
 242 GTGTGCTACCCATGATGATCACTGCTCCAAACAGCAGCTGTGTGACACCCGGGCTC 301
 995 GTTTAAGTGTCTCTGATGATGTTTGTGTGATGATGTTTGTGTGATGATGTTTGTGTG 361
 302 GTTTAAGTGTCTCTGATGATGTTTGTGTGATGATGTTTGTGTGATGATGTTTGTGTG 1114
 1055 GGATGATGCTCAGACAGAGGGGCTCAGTACATGATGATGATGATGATGATGATGATG 421
 362 GGATGATGCTCAGACAGAGGGGCTCAGTACATGATGATGATGATGATGATGATGATG 1174
 1115 AGAAGGCGACACTTGTGCTGTGCTCCGAGGCTTTACAGGGGATGATGATGATGATG 481
 422 AGAAGGCGACACTTGTGCTGTGCTCCGAGGCTTTACAGGGGATGATGATGATGATG 541
 1175 GTGTCTCCGCGACAGCCGGCTTCACTGTGTATGGGGCTGTGTGAGACATGATGTA 1234
 482 GTGTCTCCGCGACAGCCGGCTTCACTGTGTATGGGGCTGTGTGAGACATGATGTA 1294
 1235 GCTGTGTGTCAGACACCCCTGCAATATGATGATGATGATGATGATGATGATGATG 1294
 542 GTGTGTGTGTCAGACACCCCTGCAATATGATGATGATGATGATGATGATGATGATG 1346
 1295 AGAGTAT-GGTGTGTGTGTCAGACACCCCTGCAATATGATGATGATGATGATGATG 654
 602 AGAGTATGTTGT 654

RESULT 15
 BF786633 843 bp mRNA linear EST 12-JAN-2001
 LOCUS 602108585F1 NCI_CGAP_K1414 Mus musculus cDNA clone IMAGE:4236788
 DEFINITION 5', mRNA sequence.
 ACCESSION BF786633
 VERSION BF786633.1 GI:12091669
 KEYWORDS
 SOURCE house mouse.
 ORGANISM Mus musculus
 Mammalia: Eutheria: Rodentia: Sclurognathi: Muridae: Murinae: Mus.
 1 (bases 1 to 843)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: rgs@biml.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 cDNA Library Array: Incyte Genomics, Inc.
 DNA Sequencing by: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM9845 row: 3 column: 21
 High quality sequence stop: 710.
 Location/Qualifiers
 1..843
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:4236788"
 /clone_id="NCI_CGAP_K1414"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: Kidney; Vector: pCMV-SPORT6; Site: 1; NotI;
 /note="2; SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.75 kb. Constructed by Life

BASE COUNT 188 a 207 c 275 g 172 t 1 others
 ORIGIN
 Query Match 37.4%; Score 504; DB 12; Length 843;
 Best Local Similarity 97.9%; Pred. No. 2.6e-139;
 Matches 521; Conservative 0; Mismatches 10; Indels 1; Gaps 1;
 816 GCGAGACGGTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 168
 109 GCGAGACGGTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 228
 876 ACGTGTCTCCGCGACAGCCGGCTTCACTGTGTATGGGGCTGTGTGAGACATGATG 935
 169 ACGTGTCTCCGCGACAGCCGGCTTCACTGTGTATGGGGCTGTGTGAGACATGATG 995
 936 TGTGCTACCCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1055
 229 TGTGCTACCCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1115
 996 TTTAAGTGTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1175
 289 TTTAAGTGTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1234
 1056 GATGATGCTCAGACAGAGGGGCTCAGTACATGATGATGATGATGATGATGATG 1294
 349 GATGATGCTCAGACAGAGGGGCTCAGTACATGATGATGATGATGATGATGATG 1346
 1116 GAAGGCGACACTTGTGCTGTGCTCCGAGGCTTTACAGGGGATGATGATGATG 654
 409 GAAGGCGACACTTGTGCTGTGCTCCGAGGCTTTACAGGGGATGATGATGATG 654
 1176 TGTGCTCCGCGACAGCCGGCTTCACTGTGTATGGGGCTGTGTGAGACATGATG 1234
 469 TGTGCTCCGCGACAGCCGGCTTCACTGTGTATGGGGCTGTGTGAGACATGATG 1294
 1236 CTGTGTGTGTCAGACACCCCTGCAATATGATGATGATGATGATGATGATGATG 1346
 529 CTGTGTGTGTCAGACACCCCTGCAATATGATGATGATGATGATGATGATGATG 1346
 1295 AGAGTATGTTGT 640
 589 AGAGTATGTTGT 640

Search completed: February 17, 2003, 16:06:52
 Job time: 1740.73 secs

Wed Feb 19 14:09:13 2003

us-09-605-042a-1_copy_4000_6000.lst

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model
February 17, 2003, 08:33:57 ; Search time 2557.71 seconds
(without alignment)
12670.391 Million cell updates/sec

Run on:
February 17, 2003, 08:33:57 ; Search time 2557.71 seconds
(without alignment)
12670.391 Million cell updates/sec

Title: US-09-605-042a-1_COPY_4000_6000
Sequence: 1 aaatactctgttataca.....tgtatataatacataca 2001

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: em_estba:*
2: em_estbm:*
3: em_estbn:*
4: em_estbv:*
5: em_estbp:*
6: em_estbr:*
7: em_estro:*
8: em_estr:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: gb_estfun:*
15: em_estom:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	74.8	394	CNS04BCZ	AL283004 Tetraodon
2	74.8	415	AA1466217	AA1466217 vrb3c11.y
3	74.8	452	AA762356	AA762356 2M0156105
4	73.8	627	AZ853376	AZ853376 2M0249B14
5	72.2	517	AZ974581	AZ974581 2M0249B14
6	71.8	547	AZ306702	AZ306702 1M0007F19

Result No.	Score	Query Match Length	ID	Description
7	70	1173	AL514757	AL514757 Drosophila
8	69.2	987	CNS00418	CNS00418 Drosophila
9	67.6	577	AZ438500	AZ438500 1M0228E10
10	67.2	573	BH009363	BH009363 ef27910.x
11	67.2	700	AZ474721	AZ474721 1M0292L05
12	67.2	618	BH039351	BH039351 Tetraodon
13	67.2	618	BH039351	BH039351 Tetraodon
14	67.2	618	BH039351	BH039351 Tetraodon
15	66.8	618	BH039351	BH039351 Tetraodon
16	66.6	618	BH039351	BH039351 Tetraodon
17	66.2	618	BH039351	BH039351 Tetraodon
18	66.2	618	BH039351	BH039351 Tetraodon
19	66.2	618	BH039351	BH039351 Tetraodon
20	66.2	618	BH039351	BH039351 Tetraodon
21	66.2	618	BH039351	BH039351 Tetraodon
22	66.2	618	BH039351	BH039351 Tetraodon
23	65.8	618	BH039351	BH039351 Tetraodon
24	65.4	618	BH039351	BH039351 Tetraodon
25	65.2	618	BH039351	BH039351 Tetraodon
26	64.8	618	BH039351	BH039351 Tetraodon
27	64.8	618	BH039351	BH039351 Tetraodon
28	64.6	618	BH039351	BH039351 Tetraodon
29	64.6	618	BH039351	BH039351 Tetraodon
30	64.6	618	BH039351	BH039351 Tetraodon
31	64.2	618	BH039351	BH039351 Tetraodon
32	64.2	618	BH039351	BH039351 Tetraodon
33	64.2	618	BH039351	BH039351 Tetraodon
34	64.2	618	BH039351	BH039351 Tetraodon
35	63.8	618	BH039351	BH039351 Tetraodon
36	63.8	618	BH039351	BH039351 Tetraodon
37	63.4	618	BH039351	BH039351 Tetraodon
38	63.4	618	BH039351	BH039351 Tetraodon
39	63.4	618	BH039351	BH039351 Tetraodon
40	63.4	618	BH039351	BH039351 Tetraodon
41	63.2	618	BH039351	BH039351 Tetraodon
42	63.2	618	BH039351	BH039351 Tetraodon
43	63.2	618	BH039351	BH039351 Tetraodon
44	63.2	618	BH039351	BH039351 Tetraodon
45	63.2	618	BH039351	BH039351 Tetraodon

ALIGNMENTS

RESULT 1
CNS04BCZ/C
LOCUS
DEFINITION
097017 of library G from Tetraodon nigroviridis, genomic survey
sequence.
AL283004.1 GI:8021359
GSS: genome survey sequence.
Tetraodon nigroviridis.
Tetraodon nigroviridis.
Tetraodon nigroviridis.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorphi; Acanthopterygii; Percomorphi; Tetraodontiformes;
Acanthodontidae; Tetraodon.
1 (bases 1 to 394)
Jailon, O., Dasilva, C., Bouneau, L., Fisher, C.,
Roest-Crollius, H., Winkler, P., Brothier, P., Quetier, F.,
Bernot, A., Fizes, C., Winkler, P., Brothier, P., Quetier, F.,
Saurin, W., and Weissenbach, J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished
2 (bases 1 to 394)
Jailon, O., Dasilva, C., Fizes, C., Fisher, C.,
Roest-Crollius, H., Winkler, P., Brothier, P., Quetier, F.,
Bernot, A., Fizes, C., Winkler, P., Brothier, P., Quetier, F.,
Saurin, W., and Weissenbach, J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Unpublished

Wed Feb 19 14:09:13 2003

us-09-605-042a-1_copy_4000_6000.rst

Tel: 314 286 1800
Fax: 314 286 1810
Email: mouse@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
IMAGE Consortium (info@image.lnl.gov)
MGI:654604
putative full length read
vector to vector length is 544
vector: -28ml3 rev1 ET from Amersham
Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 445.
Location/Qualifiers

FEATURES
source
1. 452
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone_lib="Stratagene mouse skin (#937313)"
/clone_lib="Stratagene mouse skin (#937313)"
/sex="females"
/tissue_type="whole skin"
/dev_stage="11 weeks old"
/lab_host="SOLR (kanamycin resistant) SK-1; Site: 1; EcORI
/note="Organ: skin; Vector: pBluescript SK-1; Primer: Oligo
/Site: 2: XhoI; Cloned unidirectionally; female mice.
Average insert size: 1.0 kb; Uni-ZAP XR vector: -5'
Adaptor sequence: 5' GATTCGCGACGAG 3' -3' adaptor
sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3'

BASE COUNT 108 a 90 c 133 g 121 t

ORIGIN
Query Match 3.7%; Score 74; DB 9; Length 452;
Best Local Similarity 60.4%; Pred. No. 9.5e-08; Indels 1; Gaps 1;
Matches 139; Conservative 0; Mismatches 90;

OY 1773 TACTGTGGCAGCTTAAGTGTCTGAATCACTGATGACATGACCTTGA 1832
DB 324 TAGAGCTTTATGATGCTGAGCTGAGCTGACCTGATGATGATGATG 1891
OY 1833 GCTTCTGATCTTATATCA-CTCTCAAGTGGCCAGATTAATGATG 205
DB 264 TCTCATTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1951
OY 1892 CTCAGTTATGCTGTGCTGAGAGTAACTTATCAACACACACAT 145
DB 204 CCGAGTTTATGAGTGTGGGAGTGAATGACACACACACACACAC 145
OY 1952 CACATACACACACACACACATATATATATATATATATATAT 2001
DB 144 CAC 95

RESULT 4
A2853376 627 bp DNA linear GSS 21-FEB-2001
LOCUS A2853376 Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC2M0156105 R, DNA sequence.

ACCESSION A2853376 GI:13041427
VERSION A2853376.1
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Mammalia; Eutheria; Barber, M., Beacorn, T., Duval, B., Hamill, C.,
Dunn, D., Aoyagi, A., Meenoud, M., Meenoud, E., von Niederhausern, A.,
Islam, H., Rose, R., Stokes, R., Tinney, A., von Niederhausern, A.,
M., Rose, R., Rose, R., Stokes, R., Tinney, A., von Niederhausern, A.,
and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts

TITLE
JOURNAL
COMMENT
University of Utah Genome Center

University of Utah
Bldg. 20 S. 2030 E., SLIC, UT
Rm. 308, Biomedical
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: adunegenetics.utah.edu
Insert length: 10000 Std Error: 0.00
Plate: 0156 row: 1 column: 05
Seq primer: CACACAGAAACGCTATGAC
Class: plasmid ends
High quality sequence stop: 627.
Location/Qualifiers

FEATURES
source

1. 627
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="UUGC2M0156105"
/clone_lib="mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pMD229; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-replicated with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD22 (911473214191AF129072.1), the vector was ligated
to inducible derivative of plasmid R1, the insert adaptors and
with adaptor complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance.

BASE COUNT 168 a 138 c 118 g 203 t

ORIGIN
Query Match 3.7%; Score 73.8; DB 17; Length 627;
Best Local Similarity 65.5%; Pred. No. 1.1e-07; Indels 57; Gaps 0;
Matches 108; Conservative 0; Mismatches 57;

OY 1723 TTTTGTACAGAGGGAGTATATATATATATATATATATATATAT 158
DB 99 TTGGGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1782
OY 1783 GCGAGCTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1842
DB 159 TGCATCTTACGCTGCTGAGTGTGTGTGTGTGTGTGTGTGTGT 1887
OY 1843 CTTATATCTACGCTGCTGAGTGTGTGTGTGTGTGTGTGTGTGT 263
DB 219 CTTCTTCTTCCACCTCAAGTTCATGATTCATGATTCATGATTC 263

RESULT 5
A2974581/c 517 bp DNA linear GSS 27-APR-2001
LOCUS A2974581/c Mouse 10kb plasmid UUGC2M library Mus musculus genomic
DEFINITION clone UUGC2M0249B14 F, DNA sequence.

ACCESSION A2974581/c GI:13845808
VERSION A2974581
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Rodentia; Sciurognathi; Muridae; Murinae; Mu.
REFERENCE
AUTHORS Mammalia; Eutheria; Barber, M., Beacorn, T., Duval, B., Hamill, C.,
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,
1 (bases 1 to 517)

us-09-605-042a-1-cof1-

University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 201
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert length: 10000 Std. error: 0.00
Plate: 0249 row: B column: 14
Seq primer: CGTGTGAAACGACGCGCAGT
Seq: plasmid ends
Class: quality sequence stop: 517.
High quality sequence stop: 517.
Location/Qualifiers

/strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="U06C2M0249B14"
 /clone_1bp="mouse 10kb plasmid U06C2M library"
 /sex="Female"
 /lab_host="E. coli strain XL10-Gold, T1-resistant, F"
 /note="Vector: pMD42uv; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory (Bar Harbor, ME) and sequenced (data not shown). The DNA

BASE COUNT	96 a	122 c
ORIGIN	DB 17;	Length 517;
	Score 72.2;	

[illegible]

RESULT 6
AZ306702

ACCESSION	GI 10344968
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	

University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 E. 20th St.,
84112, USA
Tel.: 801 585 5606
Fax: 801 585 7177
Email: ddunne@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0007 row: F column: 19
Seq primer: CACACGCAACACGTATGACC
Class: Plasmid ends
High quality sequence stop: 547.
High quality extension/Qualifiers

```

"organism":"C57BL/6J"
"strain_refs":["taxon:10090"]
"db_xref":["taxon:10090"]
"clone_id":"Mouse 10Kb Plasmid UNGC1M library"
"sex":"Male"
"note":"Vector: PWD42ny, Purified genomic DNA from M. musculus C57BL/6J (male) resource"

```

[illegible]

Query Match	3.6%	Score 71.8	DB	Indels	Gaps
Best Local Similarity	68.0%	Pred. No. 3.5e-07			
Matches 100: Conservative	0	Mismatches 47			
QY 1855	CTCTCAAGTGGCCAGATTATAGTGTGCACACCTATACACGATTATGCTGTGCTAAGGA	1914			
DB 139	CTCTCAAGTGTGGATTAAAGGATGCGGCACCTCCACCTCACTATATGTTAGTTT	1974			
QY 1915	CTAAGCCCAATTATACAAACACACACATATATACACACATACACACACACACACGT	258			
DB 199	TTATATATAAACAACACACACATATATATATATATATATATACACACATACACACAT				

Db


```

OY      1975 ATATATATGATATATATATACATACA 2001
          |||||  ||  ||||| ||||| |||||
DB      259 ATATACATACATATATATATACATACA

```

RESUULT 7	1173 bp	mRNA	linear	EST 13-FEB-2001
AL514757				
LOCUS				
DEFINITION	AL514757 U1L_NF1006.PL2 Homo sapiens CDNA clone CLO8B0152D01 3			
ACCESSION	AL514757			
VERSION	AL514757			
KEYWORDS	prime, mRNA sequence.			
SOURCE	AL514757.1 GI:12778250			
ORGANISM	EST.			
	human.			
	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE	1 (bases 1 to 1173)			
AUTHORS	L., W. B., Gruber, C., Jeesse, J. and Polayes, D.			
TITLE	Full-length cDNA libraries and normalization			
JOURNAL	Unpublished (2001)			
COMMENT	Contact: Genoscope			
	Genoscope - Centre National de Sequencage			
	066 EVRY cedex - France			
	Web : www.genoscope.cns.fr .			

DocId:31173
 1. .1173
 /organism="Homo sapiens"
 /taxon.9606"
 /db_xref="taxon.9606"
 /clone="Cl0B0152D01"
 /clone_1lib="L11.NFL006_PL2"
 /tissue_type="placenta"
 /vector="pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
 /note="vector: a NotI-oligo (dN) primer. Five prime end
 was primed with a NotI-oligo (dN) primer. Five prime end
 was enriched, double-stranded cDNA was digested with Not I and
 Eco RV sites of the pCMVSPORT 6
 cloned into the Not I and Eco RV sites of the pCMVSPORT 6
 vector. Library was normalized. Library Life Technologies
 Life Technologies. Contact : Feng Liang Life Technologies
 a division of Invitrogen 9800 Medical Center Drive
 Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
 Email : liang@lifetech.com URL :
<http://fulllength.invitrogen.com>"
 98 g 160 t 69 others

BASE COUNT	DB 9;	length 1173;
ORIGIN	score 70;	
522 a		

Query Match	Score	Pred. No.	Indels	Gaps
Best Local Similarity	44.7%	240	15	1
Best Local Similarity	44.7%	240	15	1
Matches	207	Conservative		

[illegible][illegible][illegible]

```

FEATURES
source
Location="
1. .987
/organism="Drosophila melanogaster"
7337"

```

BASE COUNT	238 a	Length 987
NR 17:		

Query Match	Score	Indels	Gaps
Query Match	3.58;	Score 69.27	0
Best Local Similarity	25.7%;	Pred. No. 1.8e-06;	
Best Local Similarity	127;	Mismatches 200;	
Conservative			

[illegible]

Db 729 GAAAAAARAAAARARARARAGRGGAGAGAGCGRGRGRCGRGRCGRGRCGRARRR 670
QY 558 GTGGATCTACTGTAGAGTATGGCTGTGTGACCAAAATGTGGCAAAATGTCATCC 617
Db 669 GAGRRRRRAGRGRRRRGAGRAGRGRGRGRRAGARRRRRAGRRAGAGARGRGR 610
QY 618 TGACCAAGACAGCTTGCTGTCTTTGCCAGATCCTTAATAAAGCCACATGCCATGCATG 677
Db 609 RRRRRRRAGARGARARRRARRRRRAGARRAARAAARRGGARRGGAGRGGRGR 550
QY 678 AAGCTAGGGAATGTAGGGGAAAGGTATATGATGATGACGAAGTACCGAGAGACCAG 737
Db 549 RCGRRGGGGGCGRRGRGRCGRRRRRRGGRRRRRAGARRRRGGRRRAGARRAARAA 490
QY 738 GAAGACAGAGATGAGGAGG 757
Db 489 GAAGAGARAGRRRRRRGR 470

[illegible]

DEFINITION	ACCESSION
clone U06C1M022020	A2438500
	GI:10562513

[illegible]

Eumalacostraca: Eutheria: Rodentia: Mammalia: 1 (bases 1 to 577)
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,
 Malmoud, M., Meenen, E., Pedersen, T., Reilly, A.,
 Longacre, S., Stokes, R., Tinney, A., von Niederhausern, A.

TITLE	JOURNAL	COMMENT
Mouse whole genome scaffolding with paired end reads from 10xK and Wright.D., Weiss, R. and Wright.D., Weiss, R. (2000)	Unpublished (2000)	Contact: Robert B. Weiss University of Utah Genome Center 1500 E. 1500 S. 2030 E., SLC, UT 84143

```

FEATURES
source
    University of Utah
    Rm. 308, Biomedical Polymers Research Bldg.
    84112, USA
    Tel: 801 585 5606
    Fax: 801 585 7177
    Email: ddunn@genetics.utah.edu
    Insert Length: 10000 Std Error: 0.00
    Insert Length: 10000 Std Error: 10
    Plate: 0228 iow: E column: 10
    Seq primer: CACACAGGAAACACGATGACC
    Seq: plasmid ends
    Class: sequence stop: 577.
    High quality sequence stop: 577.
    Location/Qualifiers
        1..577
        /Organism="Mus musculus"
        /db_xref="GenBank:U00001"

```

```

/strain="CS/507/10090"
/db.xref="taxon:10028810"
/clone.lib="Mouse 10kb plasmid UUC1M library"
/sex="Male"
/lab.host="E. Coli strain XL10-Gold, T1-resistant, F-"
/lab.vector="PMD42mv; Purified genomic DNA from M.
/note="C57BL/6J (male) was obtained from the Jackson
laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative

```

[illegible]

495. CA 496
Db

[illegible]

LOCUS	ef27g10.x1	101008	2-
DEFINITION	sequence.		
ACCESSION	BH009363		
VERSION	BH009363.1	GI:13954992	

[illegible]

TITLE
JOURNAL
COMMENT

B., Nascimento, L., Bal, H., Deanda, A. oleracea
W.L.M., Zutavern, F. Shotgun Reads from Brassica
Whole Genome Shotgun
Unpublished (2001)
Contact: W. Richard McCombe
Lila Annenberg Hazen Genome Sequencing Center
Cold Spring Harbor Laboratory
PO Box 100, Cold Spring Harbor, NY 11724, USA
Tel.: 516 367 8884
Fax: 516 367 8874
Email: mcombec@cs.hl.org
Plate: ef27 row: g column: 10
Seq primer: -21fwdUniv
Class: shotgun
High quality sequence stop: 573.
Location/Qualifiers

```

FEATURES
SOURCE
1. .573
/organism="Brassica oleracea"
/ab_xref="taxon:3712"
/clone="eif2/g11"
/clone_1b="101000"
/note="Vector: M13 for .x reads, pzero-2 for .b and .g
reads; Site_1: rcory; DNA prepared as whole genome plasmid
library from young, green leaves"
library_1: from young, green leaves"
library_2: from young, green leaves"
DNAS: DNA provided by Dr. Tom Osborn, University of
Wisconsin-Madison, Department of Agronomy."

```

purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-gold (Stratagene) cells and selected for ampicillin resistance."									
BASE COUNT	253	a	120	c	106	g	219	t	2 others
ORIGIN									
Query Match	3.4%	Score	67.2	DB	17	Length	700		
Best Local Similarity	77.9%	Pred. No.	5.3e-06						
Matches	81	Conservative	0	Mismatches	23	Indels	0	Gaps	0
OY	1895	AGTTTATGCGGCTAAGAGACTAAGCCAAATTATCAACACACACATATATACACAC	1954						
Db	597	AATTACCTGTGTCTAATTCCTAAGTTTACACACACACACACACACACACACACAC	656						
OY	1955	ATACACACACACACACACGTATATATATATATATATATATATATATACAT	1998						
Db	657	ACACACACACACACACATATATATATATATATATATATATATATATAT	700						
RESULT 12									
LOCUS	BH093951	618	bp	DNA	linear	GSS	18-JUL-2001		
DEFINITION	RPci-24-297122.TV	RPci-24	Mus musculus	genomic	clone	RPci-24-297122			
ACCESSION	BH093951								
VERSION	BH093951.1	GI:14913856							
KEYWORDS	GSS.								
SOURCE	house mouse.								
ORGANISM	Mus musculus								
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.								
TITLE	1 (bases 1 to 618)								
JOURNAL	Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akınret,B., Levins,M.,								
COMMENT	Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,								
	Russell,D., de Jong,P. and Fraser,C.M.								
	Mouse BAC End Sequences from Library RPci-24								
	Unpublished (1999)								
	Other_GSSs: RPci-24-297122.TV								
	Contact: Shaying Zhao								
	Department of Eukaryotic Genomics								
	The Institute for Genomic Research								
	9712 Medical Center Dr., Rockville, MD 20850, USA								
	Tel: 301 838 0200								
	Fax: 301 838 0208								
	Email: szhao@tigr.org								
	Clones are derived from the mouse BAC library RPci-24. For BAC								
	library availability, please contact Pieter de Jong								
	(pdejong@mail.cho.org). Clones may be purchased from BACPAC								
	Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end								
	Page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html								
	Plate: 297 row: I column: 22								
	Seq primer: T7								
	Class: BAC ends.								
FEATURES	Location/Qualifiers								
Source	1..618								
	/organism="Mus musculus"								
	/strain="C57BL/6J"								
	/db_xref="taxon:10090"								
	/clone="RPci-24-297122"								
	/clone_lib="RPci-24"								
	/sex="Male"								
	/cell_type="Spleen/Brain"								
	/note="Vector: pTARBAcl; Site_1: BamHI; Site_2: BamHI;								
	RPci-24 Mouse BAC library produced by Pieter de Jong. The								
	library was cloned in the pTARBAcl cloning vector at the								
	BamHI sites using MboI partially digested male C57BL/6J								
	DNA."								
BASE COUNT	202	a	146	c	129	g	141	t	
ORIGIN									
Query Match	3.3%	Score	67	DB	17	Length	618		
Best Local Similarity	67.6%	Pred. No.	5.9e-06						

Matches	94: Conservative	0: Mismatches	45: Indels	0: Gaps	0:
QY 1516	CATGGCATGTTTTTTTTTGTATGAAACACACAGCAGCAGCAGCATGCTGCGC	1575			
DB 295	CTTAGCGCTTCTCAGGGGTGACATAAAGCAGCATGTCAGCGGACACACATACAC	354			
QY 1576	AGCGGGCACACACACAGCGGCACACACACACACACACCATGACACATGACACACA	1635			
DB 355	ACACACACACACACACACACACACACACACACACACACAAACATATACCCAAA	414			
QY 1636	CAACTGCAAAAGTGAATA 1654				
DB 415	ACACACACACAAATATAAA 433				
RESULT 13					
CNS02DLP		762 bp	DNA	linear	GSS 13-MAY-2000
LOCUS	Tetraodon nigroviridis genome survey sequence T7 end of clone				
DEFINITION	26001 of library G from Tetraodon nigroviridis, genomic survey				
ACCESSION	AL192598				
VERSION	AL192598.1				
KEYWORDS	GSS: genome survey sequence.				
SOURCE	Tetraodon nigroviridis.				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.				
REFERENCE	1 (bases 1 to 762)				
AUTHORS	Roest-Crolius,H., Jalllon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Winkler,P., Brotlier,P., Quetier,F., Saurin,W. and Weissenbach,J.				
TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 762)				
AUTHORS	Roest-Crolius,H., Jalllon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.				
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis				
JOURNAL	Unpublished				
REFERENCE	3 (bases 1 to 762)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (12-APR-2000)				
COMMENT	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/tetraodon .				
FEATURES	Location/Qualifiers				
SOURCE	1..762				
	/organism="Tetraodon nigroviridis"				
	/db_xref="taxon:99883"				
	/clone="260J01"				
	/clone_1bp="G"				
	/notes="Genoscope sequence ID : C0AG260CE01LP1-end : T7"				
BASE COUNT	256 a 170 c 163 g 154 t 19 others				
ORIGIN					
Query Match	3.3%: Score 67; DB 17; Length 762;				
Best Local Similarity	59.3%: Pred. No. 6.1e-06;				
Matches	99: Conservative 8; Mismatches 60; Indels 0; Gaps 0;				
QY 1540	AAACACACGACAGCAGCAGCAGCTACGCTGTCGACAGCGGCACACACACAGCGC	1599			
DB 286	AC	345			
QY 1600	ACACACACACACACAGCATGACACATGACACACACACAAACTGCAAAAGTGAATA	1659			
DB 346	ACACACACACACACAGCGCGCGCATCCAGCAACATTTCCATVAAAAA	405			

QY	1660	ATATTTCTCATCTTGGCAAAAGTGATGGAAATGTATCAAAAAGGAA	1706
Db	406	AAAMGHTTCTGGGSGTGGGGGGAGCCMRAGTMTCTGGAGATGAA	452
RESULT 14			
CNS007JJ			
LOCUS			
DEFINITION		CNS007JJ 1203 bp DNA linear GSS 03-JUN-1999	
ACCESSION		Drosophila melanogaster genome survey sequence TET3 end of BAC #	
VERSION		BACR15F04 of RPCI-98 library from Drosophila melanogaster (fruit	
KEYWORDS		fly), genomic survey sequence.	
SOURCE		AL067347	
ORGANISM		AL067347.1 GI:4945810	
		GSS.	
		Drosophila melanogaster.	
		Drosophila melanogaster	
		Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;	
		Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
		Ephydroidea; Drosophilidae; Drosophila.	
		1 (bases 1 to 1203)	
REFERENCE		Genoscope.	
AUTHORS		Direct Submission	
TITLE		Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage ;	
JOURNAL		Bp 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr	
COMMENT		- Web : www.genoscope.cns.fr)	
		determination of this BAC-end sequence was carried out as part of a	
		collaboration with the Berkeley Drosophila Genome Project (BDGP).	
		The BDGP is constructing a physical map of the Drosophila	
		melanogaster genome using these BACs. For further information	
		please see http://www.fruitfly.org/TheBDGP/Drosophila	
		melanogaster BAC library was prepared by Kazutoyo Osoegawa and	
		Aaron Mammeter in Pieter de Jong's laboratory in the Department of	
		Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,	
		NY. The library is named RPCI-98 and was constructed by partial	
		Ecotri digestion of Drosophila DNA provided by the BDGP from the	
		isogenic strain v2; cn bw sp, the same strain used for the BDGP's	
		PL and EST libraries. A more detailed description of the library	
		and how to order individual BAC clones, the entire library, or	
		filters for hybridization from the BACPAC Resource Center can be	
		found at http://bacpac.med.buffalo.edu/drosophila_bac.htm .	
FEATURES		Location/Qualifiers	
source		1.1203	
		/organism="Drosophila melanogaster"	
		/db_xref="taxon:7227"	
		/clone="BACR15F04"	
		/clone_lib="RPCI-98"	
		/note="end : TET3"	
BASE COUNT		483 a 310 c 52 g 174 t 184 others	
ORIGIN			
Query Match		3.38; Score 67; DB 17; Length 1203;	
Best Local Similarity		44.7%; Pred. No. 6.6e-06;	
Matches 210; Conservative		9; Mismatch 243; Indels 8; Gaps 2;	
QY	1540	AAACACACGACGACGGCAGGACACTCAGCTGTCGGCCAGCGACACACACGCGC	1599
Db	289	ACACACACACACACACACACATATACGCTACACACACACACACACACACAC	348
QY	1600	ACACACACACACACGCGATGACACATGCACACACACAAACTGCAAAAGTGAATAAAG	1659
Db	349	ACACACATACACACACACACACACACACATGANGNGNGNACACACACACACAC	408
QY	1660	ATATTTCTCAGTTGGCAAGTGTATGAAGTTCATTAATTAAGTAAGTATTAATCTAGAA	1719
Db	409	ACACACATGTNNATATATATACACACACACACATACACACATACACACGTCGCG	468
QY	1720	CTATTTTGACATGAGGAGGAGTTATTAATATATTTGTTATTTGTTATTTTACTGTT	1779
Db	469	ACACACACATATATACGCGCGACACACACATATATATACACATATATATATATACATAC	528
QY	1780	TGTGGCAGCTTAAGTGTGCTTGAAGTCACTATGAAGTAGCAATG-ACCTTGAGCTTCT	1838

us-09-605-042a-1-copy-4000-6000.rst

[illegible]

Search completed: February 17, 2003, 16:06:32
Job time : 2578.96 secs

1

us-09-605-042a-1_copy-1-1500.rst

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd

OM nucleic - nucleic search, using sw model
February 17, 2003, 08:33:57 ; Search time 1917.32 Seconds
(without alignments)
12670.391 Million cell updates/sec

Title: US-09-605-042A-1_COPY-1-1500
 Perfect score: 1500
 Sequence: 1 gggggggccctcggagttt.....agatcccccattctctctcg 1500

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 32308132
Chosen parameters:

Total number of hills seen: 0

Minimum	DB	seq	Length:	2000000000
Maximum	DB	seq		

post-processing:	Maximum	Match	100%
	Maximum	first 45 summaries	
	Maximum	first 45 summaries	

```

Database :
EST: *
1: em_estba: *
    +hum: *

```

```

1: em_estba:*
2: em_esthnm:*
3: em_estlm:*
4: em_estlmv:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hnc:*
9: gb_estl:*
10: gb_estc:*
11: gb_hnc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: gb_gss_hum:*
18: em_gss_hum:*
19: em_gss_liv:*
20: em_gss_pln:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

```

27: en_lgs_tod:
pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the total score distribution.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	59.2	3.9	997	17	CNS005TE	AL060637 Drosophill
2	55.8	3.7	987	17	CNS004I8	AL066537 Drosophill
3	55.6	3.7	1101	17	CNS0039G	AL063821 Drosophill
4	55	3.7	1101	17	CNS0100X	AL068379 Drosophill
5	50.8	3.4	1101	17	CNS0106X	AL098555 Drosophill
6	49.6	3.3	1101	17	CNS000D1	AL065414 Drosophill

[illegible]

ALIGNMENTS

RESULT 1		DNA linear	end of BAC #
CNS005TE	997 bp		
LOCUS	CNS005TE		
DEFINITION	Drosophila melanogaster genome survey sequence TE13 end of BAC #		
	Drosophila melanogaster genome survey sequence.		
	BACR12K2 of RPCI-98 library from Drosophila melanogaster (fruit		
	fly) survey sequence.		

ACCESSION	AL060767.1	GI:494373
VERSION	AL060767.1	
KEYWORDS	GSS.	
SOURCE	Drosophila melanogaster.	
ORGANISM	Drosophila melanogaster.	
	Phylum: Arthropoda; Hexapoda; Insecta; Pterygota.	
	Class: Insecta; Pterygota; Diptera; Brachycera; Muscomorpha.	
	Order: Pterygota; Diptera; Brachycera; Muscomorpha.	
	Family: Pterygota; Diptera; Brachycera; Muscomorpha.	
	Genus: Pterygota; Diptera; Brachycera; Muscomorpha.	
	Species: Pterygota; Diptera; Brachycera; Muscomorpha.	

Epiphyas 1 to 997
1 (bases

REFERENCE
1 (bacc)
Genoscope - Centre National de Séquençage
Auteurs Direct Submission Genoscope - Centre National de Séquençage
Titre Direct Submission Genoscope (E-mail : segretef@genoscope.cns.fr)
Journal Submitted (02-JUN-1999)
No 191 91006 EVRY cedex - FRANCE
Carried out as part of the GENOSCOPE project
© 1999 Centre National de Séquençage

COMMENT

Submitted 1906. ENVR cedex - France). The sequence was carried out as part of the BAC-based genome project (BDGP Web: www.genoscope.cns.fr). Determination of this BAC-end sequence was carried out in collaboration with the Berkeley Drosophila Genome Project (BDGP Web: <http://www.fruitfly.org>). The BDGP is constructing a physical map of the Drosophila genome using these BACs. For further information please see <http://www.fruitfly.org>. The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Gotohawa and Aaron Kammmerer in Peter de Jong's laboratory in the Buffalo Cancer Genetics at the Roswell Park Cancer Institute by partial digestion of the Roswell Park Cancer Institute (RPCI-98) library. The library is named RPCI-98 and was constructed by partial

Wed Feb 19 14:09:07 2003

us-09-605-042a-1-copy_1_1500.rst

ECORI digestion of drosophila DNA provided by the BDGP from the isogenic strain Y2; on bw sp, the same strain used for the library and EST libraries. A more detailed description of the library, or pl and EST libraries, the entire library, can be found at <http://bacpac.med.buffalo.edu/drosophila.bac.htm>.

FEATURES
source
1. .987
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCL-98"
/note="end : TET3"

BASE COUNT 89 a 99 c 258 t 538 others
ORIGIN 3.9% Score 59.2; DB 17; Length 997;
Query Match 20.1% Pred. No. 1.4e-05; Indels 2; Gaps 2;
Best Local Similarity 189; Mismatches 358;

64 ATGAGTCCAGTCAAGCTCATCTGATGCTGATACCAATAGTACCATTAT 183
312 AT 431
124 ATGATCAAGCCGAGCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTG 243
372 ANAAAAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 491
184 CATTATGATTTTACCATTTCCCTCCATGAGGAGGAGGAGGAGGAGG 303
432 NNANANNNNAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 551
244 CTACTATGCTGATGATGATGATGATGATGATGATGATGATGATGAT 611
492 TCTTTCCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 670
304 ATGATCAAGCTCATCTGATGATGATGATGATGATGATGATGATGAT 730
552 TTTTTCCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 790
364 TCAAGTTCCTGATGATGATGATGATGATGATGATGATGATGATGAT 850
612 TTTTTCCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 910
424 TTTTTCCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 970
671 TCTTTCCTGATGATGATGATGATGATGATGATGATGATGATGATGAT
484 AGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
731 TTTTTCCTGATGATGATGATGATGATGATGATGATGATGATGATGAT
544 TTTTTCCTGATGATGATGATGATGATGATGATGATGATGATGATGAT
791 TTTTTCCTGATGATGATGATGATGATGATGATGATGATGATGATGAT
604 CTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
851 TCTTTCCTGATGATGATGATGATGATGATGATGATGATGATGATGAT
663 CAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
911 TCTTTCCTGATGATGATGATGATGATGATGATGATGATGATGATGAT
723 TGCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
971 TCTTTCCTGATGATGATGATGATGATGATGATGATGATGATGATGAT

RESULT 2 987 bp DNA linear GSS 03-JUN-1999
CNS00418 Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage
BP 191 91000 Evry cedex - FRANCE (E-mail : segregen@genoscope.cns.fr)
- Web : www.genoscope.cns.fr
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamoser in the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCL-98 and was constructed from the
ECORI digestion of drosophila DNA provided by the BDGP's
isogenic strain Y2; on bw sp, the same strain used for the library,
pl and EST libraries. A more detailed description of the library, or
filters for hybridization from the BACPAC Resource Center can be
found at <http://bacpac.med.buffalo.edu/drosophila.bac.htm>.

FEATURES
source
1. .987
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCL-98"
/note="end : TET3"

BASE COUNT 238 a 162 c 177 t 393 others
ORIGIN 3.7% Score 55.8; DB 17; Length 987;
Query Match 23.5% Pred. No. 0.00013; Indels 1; Gaps 1;
Best Local Similarity 154; Mismatches 251;

195 TTACCAATTCCTGATGATGATGATGATGATGATGATGATGATGATGAT 314
448 TTTTTCCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 567
255 CTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 627
508 TTTTTCCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 687
315 CTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 747
568 TTTTTCCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 806
375 TGAGAGCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 866
628 TTTTTCCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 926
435 ATGATCAAGCTCATCTGATGATGATGATGATGATGATGATGATGAT 986
688 TTTTTCCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1046
495 ATGATCAAGCTCATCTGATGATGATGATGATGATGATGATGATGAT 1106
748 TTTTTCCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1166
555 CTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1226
807 TTTTTCCTGATGATGATGATGATGATGATGATGATGATGATGATGAT

OY	631	ATTCASCTGTTGAATGTCTTGAACAAATAAATAATTCTGCAGGGTGTTGCTC	690
Db	715	TNNYAHUYUUYIMAYUMIWTCTACTYNNNNNYHMYNTTUAMNMMNNANRA	774
OY	691	TASSCAATTCAGCCSTTCAGAGTAATACTGCC--CATCTTCATCTCTGCT	747
Db	775	AAMAATAATNNHYNTHTNNHTYMUHXYMUTSCSYTHCMNHNTAPCWITNNHM	834
OY	748	TTCGTCTTCAATCTCAATATGTAAGCCATTTCTAAAAGCTCTCCAAGACAATATT	807
Db	835	TWNHWNNHTWNNHTTNHAMNHTNCMMWWMTATWTMAATNCACMTMNHNMMNNNN	894
OY	808	GAAAGACTTTGCGATCTCATATGATGATACATACATATGCTGCTGCCACASCATGA	867
Db	895	MCHANNHTMCNMNNNNHCNNHNTYTUNMTCSNMNNNNNNMMATMTITTMIM	954
OY	868	CCATCCCCATGAATACAGACATGCGCTTCTTAGTCTTTCGTATGTCTGTGTA	927
Db	955	MSCMHNNCHMYNNHMYMUSCHYUCTGHATATYHNYCTGYNHCTYHNTYUAWMT	1014
OY	928	CATTGTAGATTAATGCTGTATAAACAT	955
Db	1015	AHAATTAATWWWMMHMAHMAWTMMWWW	1042
RESULT 4			
LOCUS	CNS0100X/c		
DEFINITION	CNS0100X	1101 bp	DNA linear GSS 26-JUL-1999
ACCESSION	Drosophila melanogaster genome survey sequence SP6 end of BAC		
VERSION	BACN03G04 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.		
SOURCE	AL098379		
KEYWORDS	AL098379.1 GI:5609990		
ORGANISM	GSS:		
ORGANISM	Drosophila melanogaster.		
ORGANISM	Drosophila melanogaster		
ORGANISM	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;		
ORGANISM	Nephrolepta; Endopterygota; Diptera; Brachycera; Muscomorpha;		
ORGANISM	Ephyroidae; Drosophilidae; Drosophila.		
REFERENCE	1 (bases 1 to 1101)		
AUTHORS	Genoscope.		
TITLE	Direct Submission		
JOURNAL	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr - Web : www.genoscope.cns.fr)		
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk . This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.		
FEATURES			
SOURCE	Location/Qualifiers		
	1..1101		
	/organism="Drosophila melanogaster"		
	/db_xref="taxon:7227"		
	/clone="BACN03G04"		
	/clone_1ib="DrosBAC"		
	/plasmid="pBelobAC11"		
	/note="end : Spe"		
BASE COUNT	195 a 108 c 131 g 161 t 506 others		
ORIGIN			
Query Match	3.7%; Score 55; DB 17; Length 1101;		
Best Local Similarity	16.3%; Pred. No. 0.00024;		
Matches	Conservative 184; Mismatches 175; Indels 0; Gaps 0;		
OY	63	TATGAGTTCAGTACACAGCTATCTCTAGATGTCTGCGATCAATAAGTGACCATATCT	122
Db	1044	TNNKMWYUYNCRKYCKJUNYUUTCHTSCMUCSCWSCSMCMWHMMHMMHMMHMAVU	985

JOURNAL
COMMENT

Unpublished (1999)
Other GSSs: RPCI-24-210F14.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end page: http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html. Plate: 210 row: F column: 14
Seq primer: SP6
Class: BAC ends.

FEATURES
source

Location/Qualifiers
1. 804
/organism="Mus musculus"
/strain="C57BL/6j"
/db_xref="taxon:10090"
/clone="RPCI-24-210F14"
/clone_lib="RPCI-24"
/sex="male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI; RPCI-24 Mouse BAC library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."
BASE COUNT 195 a 176 c 236 g 197 t
ORIGIN

Query Match 3.2%; Score 47.6; DB 17; Length 804;
Best Local Similarity 69.1%; Pred. No. 0.028; 29; Indels 0; Gaps 0;
Matches 65; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 1172 TGATGCACACCTGTAATTCGTGCTAGGAGACAGATGAGGGGAGCTATGATTAAG 1231
DB 657 TGCTGCAGCTCTGTAATTCACACTGCTGGAGCAGAGCAGAGCAATTTAGG 716
QY 1232 CCCATTTTAACTGCTGGGAGACACTGTTT 1265
DB 717 CCAACTTAGCAACACTTAGAGAACTGTTCT 750

RESULT 11

AQ321637 558 bp DNA linear GSS 06-MAY-1999
LOCUS RPCI11-93G5.TV RPCI-11 Homo sapiens genomic clone RPCI-11-93G5, DNA
DEFINITION sequence.
ACCESSION AQ321637
VERSION AQ321637.1 GI:4054371
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wille,C., de Jong,P. and Venter,J.C.
TITLE Use of human BAC end sequences for Sequence-Ready Map Building
JOURNAL Unpublished (1998)
COMMENT Other GSSs: RPCI11-93G5.TV
Contact: Shaying Zhao, William Niernan, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbeet@tigr.org

Clones are derived from the human BAC library RPCI-11. For BAC

library availability, please contact Pieter de Jong (pdejong@mail.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/cdb/humgen/bac_end_search/bac_end_search.html. Seq primer: T7
Class: BAC ends.

FEATURES
source

Location/Qualifiers
1. 558
/organism="Homo sapiens"
/db_xref="GDB:7535476"
/db_xref="taxon:9606"
/clone="RPCI-11-93G5"
/clone_lib="RPCI-11"
/sex="male"
/cell_type="Lymphocytes"
/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI; RPCI11 Human Male BAC library"
BASE COUNT 189 a 86 c 123 g 160 t
ORIGIN

Query Match 3.1%; Score 46.6; DB 17; Length 558;
Best Local Similarity 57.5%; Pred. No. 0.045; 74; Indels 2; Gaps 1;
Matches 103; Conservative 0; Mismatches 74; Indels 2; Gaps 1;

QY 1081 TAACACAACCTTAATAGATTAAACAGCAGGATTTATTCACATGTTTGAGACGCCA 1140
DB 503 TACCAACAACCTTGTGACTTAACACACAAATTTACTCCCTTATAGTTCTGGGGTCA 444
QY 1141 GAAATCTGACACAGTTTCAATTTAGACTGTGATGACACACCTGTAATTCGTGACTTAG 1200
DB 443 GAAGTCTGAATATGATTCTACTGGGCAAAAGTCAAGTGTCAGTTGGCTGGCTCTTCT 384
QY 1201 GAGGACAGATGACAGGAGCATGATTTAAACCCATTTTATAGCTGCTGGTGAGAAC 1259
DB 383 AGAG--GCCCAAGGGGTGATTCATTTACTTTTTCAGCTTCTAGATGATGACC 327

RESULT 12

CNS00CF2 1031 bp DNA linear GSS 04-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BAC25K05 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL059199
VERSION AL059199.1 GI:4946662
KEYWORDS GSS.
SOURCE Drosophila melanogaster.

REFERENCE Drosophila melanogaster.
AUTHORS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
TITLE 1 (bases 1 to 1031)
JOURNAL Genoscope.

COMMENT Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazuhiro Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

Wed Feb 19 14:09:07 2003

us-09-605-042a-1-copy-1-1500.rst

FEATURES
source location/Qualifiers
1.1031 /organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone.lib="RPCL-98"
/note="end: TET3"

BASE COUNT 176 a 169 c 124 g 214 t 348 others

Query Match 3.1%; Score 46.6; DB 17; Length 1031; 1;
Best Local Similarity 22.1%; Pred. No. 0.063; Indels 1; Gaps 1;
Matches 149; Conservative 198; Mismatches 326;

67 AGATCAGGAGACAGCTCTCTAGATGCTGCATGACATTAAGACCCATGATG 126
1021 ACMTTTCCHCAVCTTTCGCCCTCATMAAATAAATAAATAAATAAATAA 962
127 CAATCAGCCGACATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 186
961 AAATAAT 246
187 ATATGATTTTACATATATATATATATATATATATATATATATATATAT 843
902 ATATGATTTTACATATATATATATATATATATATATATATATATATAT 306
247 ATATGATTTTACATATATATATATATATATATATATATATATATATAT 783
842 ATATGATTTTACATATATATATATATATATATATATATATATATATAT 366
307 ATATGATTTTACATATATATATATATATATATATATATATATATATAT 723
782 ATATGATTTTACATATATATATATATATATATATATATATATATATAT 426
367 ATATGATTTTACATATATATATATATATATATATATATATATATATAT 663
722 ATATGATTTTACATATATATATATATATATATATATATATATATATAT 486
427 ATATGATTTTACATATATATATATATATATATATATATATATATATAT 603
662 ATATGATTTTACATATATATATATATATATATATATATATATATATAT 546
487 ATATGATTTTACATATATATATATATATATATATATATATATATATAT 543
602 ATATGATTTTACATATATATATATATATATATATATATATATATATAT 606
547 ATATGATTTTACATATATATATATATATATATATATATATATATATAT 483
542 ATATGATTTTACATATATATATATATATATATATATATATATATATAT 666
607 ATATGATTTTACATATATATATATATATATATATATATATATATATAT 423
482 ATATGATTTTACATATATATATATATATATATATATATATATATATAT 726
667 ATATGATTTTACATATATATATATATATATATATATATATATATATAT 363
422 ATATGATTTTACATATATATATATATATATATATATATATATATATAT 740
727 ATATGATTTTACATATATATATATATATATATATATATATATATATAT 349
362 ATATGATTTTACATATATATATATATATATATATATATATATATATAT

RESULT 13
LOCUS A0109909 458 bp DNA linear GSS 29-AUG-1998
DEFINITION C17-HSP-2378A6-TR C17-HSP Homo sapiens genomic clone 2378A6, DNA
ACCESSION A0109909
VERSION A0109909.1 GI:3486599
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Use of a random human BAC End Sequence Database for Sequence-Ready
Map Building
Unpublished (1998)
Other GSS: C17-HSP-2378A6.TF
Contact: Mark Adams
Department of Genomic Research
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 0208
Email: madams@ligr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.ligr.org/cdb/humgen/Bac_end_search/Bac_end_search.html.
Seg primer: M13 Reverse
Class: BAC ends.

FEATURES
source location/Qualifiers
1.458 /organism="Homo sapiens"
/db_xref="taxon:9606"
/clone.lib="C17-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBeloBAC11; Site:1: HindIII; Site:2:
HindIII"

BASE COUNT 166 a 92 g 128 t 2 others

Query Match 3.1%; Score 46.4; DB 17; Length 458; 0;
Best Local Similarity 48.5%; Pred. No 0.046; Indels 0; Gaps 0;
Matches 128; Conservative 0; Mismatches 136;

905 TGCGATGATGTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1024
45 TGCGATGATGTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 164
965 AAATGATTCATCGATAGACACCTCTCTGAGAGCATGATTCATGATGTTTCC 1084
105 AAATGATTCATCGATAGACACCTCTCTGAGAGCATGATTCATGATGTTTCC 224
1025 AGCATGATTCATCGATAGACACCTCTCTGAGAGCATGATTCATGATGTTT 284
165 TAAAGCAGGAGAAATAATATATATATATATATATATATATATATATATAT 1144
1085 ACAATCTTAAT 1168
225 AGATATTTTAT 308
1145 TCTGACACCATTTCAATGTTTAA 1168
285 AATTTCATGTTTACAGAAATTTG

RESULT 14
LOCUS A0109909 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION C17-HSP-2378A6-TR C17-HSP Homo sapiens genomic clone 2378A6, DNA
ACCESSION A0109909
VERSION A0109909.1 GI:4958603
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens


```

Dd      679  MMAMMMMTMMTMMMAAMTATMMMTATMTAMTMMMAAMATMMAMTAAATATTATTATWE 620
Oy      480  TTTCAGTACTACTGATCTACTGCTCTACCCGTAACTGGTATCTCATGACACTCGCA 539
Dd      619  TTTAATATTAATTTTNNNNNTTNNNTTNNNTNNNNMTNNNNNTNNNTNNNTNNNTNTNTN 560
Oy      540  ATATTTCCATTTCTCTATGCTGCAAGCTGTGAGACGTAGCTTAATGATCTCCTT 599
Dd      558  NNTTTTTTTTTTTTTTMMNTTTTTNNNNNTNNVNTGNTNNNTTNNNNNTTNNNT 500
Oy      600  TTCTCTCAGGCTCTCCGCTGCTCCACACACAT 633
Dd      499  CTCNNNTATNTCTGNTTNNMNNNNNNCNTT 466

```

Search completed: February 17, 2003, 16:06:11
Job time : 1929.57 secs

Wed Feb 19 14:09:14 2003

us-09-605-042a-1_copy_8000_9345.rge

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
February 17, 2003, 08:20:43 ; Search time 3361.36 Seconds
Run on: 11653.717 Million cell updates/sec

Title: US-09-605-042a-1_COPY_8000_9345
Sequence: 1 gacgtgatgtctgtgttaa.....gtctgcacggtgtgtaccg 1346

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: gb_da:*
2: gb_hg:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_ov:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vt:*
15: gb_vt:*
16: em_hum:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_mu:*
21: em_or:*
22: em_or:*
23: em_or:*
24: em_or:*
25: em_or:*
26: em_or:*
27: em_or:*
28: em_or:*
29: em_or:*
30: em_or:*
31: em_or:*
32: em_or:*
33: em_or:*
34: em_or:*
35: em_or:*
36: em_or:*
37: em_or:*
38: em_or:*
39: em_or:*
40: em_or:*
41: em_or:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1344.4	99.9	219386	2 AC125147	AC125147 Mus muscu
2	647.8	48.1	165503	2 AC111518	AC111518 Rattus no
3	647.8	48.1	165503	2 AC111518	AC111518 Rattus no
4	526.2	39.1	2229	10 BC010401	BC010401 Mus muscu
5	513.4	38.1	2343	10 MUSTAMK	MUSTAMK Mus muscu
6	430.2	32.0	2227	10 RATRUMOD	RATRUMOD Mus muscu
7	430.2	32.0	2227	10 RATRUMOD	RATRUMOD Mus muscu
8	311.2	23.1	168887	9 AC106796	AC106796 Homo sapi
9	283.8	21.1	2290	6 AX336081	AX336081 Sequence
10	283.8	21.1	2290	6 AX336081	AX336081 Sequence
11	283.8	21.1	2290	6 AX336081	AX336081 Sequence
12	283.8	21.1	2290	6 AX336081	AX336081 Sequence
13	283.8	21.1	2290	6 AX336081	AX336081 Sequence
14	283.8	21.1	2290	6 AX336081	AX336081 Sequence
15	256.4	19.0	2340	4 AK055722	AK055722 Homo sapi
16	169.8	12.6	2180	9 AK096043	AK096043 Homo sapi
17	104	7.7	2057	9 AF511317	AF511317 Sus scrofa
18	86	6.4	6727	4 AY061638	AY061638 Homo sapi
19	70.2	5.2	173127	9 AC021451	AC021451 Homo sapi
20	65.4	4.8	74918	9 AC004140	AC004140 Homo sapi
21	64	4.8	167031	2 AC004097	AC004097 Homo sapi
22	63.8	4.7	7053	3 ERE306453	ERE306453 Sequence
23	63.8	4.7	7053	3 ERE306453	ERE306453 Sequence
24	62.4	4.6	175648	2 AP000925	AP000925 Homo sapi
25	61	4.5	185405	2 AP001887	AP001887 Homo sapi
26	61	4.5	185405	2 AP001887	AP001887 Homo sapi
27	60.8	4.5	185405	2 AP001887	AP001887 Homo sapi
28	60.2	4.5	185405	2 AP001887	AP001887 Homo sapi
29	60.2	4.5	185405	2 AP001887	AP001887 Homo sapi
30	60.2	4.5	185405	2 AP001887	AP001887 Homo sapi
31	59.8	4.4	173976	30 AC073045	AC073045 Homo sapi
32	59.8	4.4	173976	30 AC073045	AC073045 Homo sapi
33	59.4	4.4	125372	2 AC102335	AC102335 Mus muscu
34	59.4	4.4	125372	2 AC102335	AC102335 Mus muscu
35	59.4	4.4	125372	2 AC102335	AC102335 Mus muscu
36	59.2	4.4	117872	6 AL162396	AL162396 Homo sapi
37	59	4.4	761	6 AP000390	AP000390 Homo sapi
38	59	4.4	50188	9 H3552018	H3552018 Homo sapi
39	59	4.4	59104	9 AF190641	AF190641 Homo sapi
40	59	4.4	89102	9 AF191069	AF191069 Homo sapi
41	58.8	4.4	108965	9 AP001707	AP001707 Homo sapi
42	58.8	4.4	108965	9 AP001707	AP001707 Homo sapi
43	58.8	4.4	108965	9 AP001707	AP001707 Homo sapi
44	58.6	4.4	141274	2 AF374376	AF374376 Oikopleur
45	58.6	4.4	170346	9 AC020626	AC020626 Homo sapi

ALIGNMENTS

RESULT 1
AC125147
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE

AC125147
Mus musculus chromosome UNK clone RP24-350A15, WORKING DRAFT
SEQUENCE, 8 unordered pieces.
AC125147
AC125147.2 GI:21539180
HTG: HTGS_PHASE1: HTGS_DRAFT.
house mouse
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Scleroglossa; Muridae; Murinae; Mus
1 (bases 1 to 219386)
McPherson,J.D. and Waterston,R.H.
The sequence of Mus musculus clone

Pred No is the number of results predicted by chance to have a

JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 219386)
AUTHORS	McPherson, J.D. and Waterston, R.H.
TITLE	Direct Submission
JOURNAL	Submitted (20-JUN-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE	3 (bases 1 to 219386)
AUTHORS	McPherson, J.D. and Waterston, R.H.
TITLE	Direct Submission
JOURNAL	Submitted (22-JUN-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT	On Jun 22, 2002 this sequence version replaced gi:21940647.

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wason.wustl.edu
Project Information
Center project name: M_BB0350A15
-----
Summary Statistics
-----

```

Sequencing vector: pLM2: 100%
Sequencing vector: pLM2: 0% of reads
Chemistry: Dye-terminator Big Dye: 0.990319
Chemistry: Dye-terminator Big Dye: 0.990319
Assembly program: Phrap: version 0
Consensus quality: 216121 bases at least Q40
Consensus quality: 216169 bases at least Q30
Consensus quality: 216719 bases at least Q20

Consensus quality: 171000; agarose-fp
Insert size: 171000; sum-of-contigs
Insert size: 217980; sum-of-contigs
Quality coverage: 14.92 in Q20 bases; sum-of-contigs
Quality coverage: 11.97 in Q20 bases; sum-of-contigs

* NOTE: this is a 'working draft' order of the pieces
* consists of 8 contigs. The order in this sequence record is
* is not known and their other contigs are representative
* arbitrary. Gaps between the sizes of the gaps are sequence
* runs of N, but the exact sizes of the gaps are sequence
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved. 105. Contig of 105 bp in length

	as soon as possible.	
1	105: contig of 105 bp in length	
*	205: gap of unknown length	
*	106	806: contig of 601 bp in length
*	206	906: gap of unknown length
*	807	contig of 4363 bp in length
*	907	5269: contig of unknown length
*	5270	5369: gap of 11024 bp in length
*	5370	16593: contig of unknown length
*	16394	16493: gap of unknown length
*	16494	34751: contig of 18158 bp in length
*	34652	80845: gap of unknown length
*	34752	80845: contig of 46094 bp in length
*	80846	80945: gap of unknown length
*	80946	138242: contig of 57297 bp in length
*	138243	138342: gap of unknown length
*	138243	219386: contig of 81044 bp in length
*	138343	

```
1. :219386
/organism="Mus musculus"
/db_xref="taxon:10090"
chromosome="UNK"
```

```
misc_feature 1. 105
              /note="assembly_name:Config12"
misc_feature 206. 806
              /note="assembly_name:Config14"
misc_feature 907. 569
              /note="assembly_name:Config19"
misc_feature 5370. 16393
              /note="assembly_name:Config20"
misc_feature 16494 34651
```

```

/note="assembly_name:Contig21"
34752. .80845
/note="assembly_name:Contig22"
80946. .138242
/note="assembly_name:Contig23"
138343. .219386
/note="assembly_name:Contig24"
60953 a 47433 c 47470 g 62799 t 731 others
BASE COUNT
ORIGIN
00.9%: Score 1344.4; DB 2; Length 219386;

```

[illegible]

us-09-605-042a-1_copy-8000_9345.rge

[illegible][illegible]

Mammalia: Eutheria, Rodentia, Sciurognathi, Muridae: Murinae;
REFERENCE
AUTHORS
1 (bases 1 to 187332)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-oman,F.R., Allen,C.,
Alstbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbieri,L.J., Benton,J., Bimge,K., Blankenburg,K., Bonnin,D.,
Boucek,J., Bowe,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dedecich,D.A.,
Dehne,K.R., Delgado,O., Dean,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Franz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homsi,F., Howard,S., Huber,J., Huylk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Kovach,J., Kovar,C.,
Kratovic,U., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louised,H.,
Locado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maleshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., Mcleod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,S., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokwkw,S., Ogih,M., Okumou,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peeters,L., Plickens,R., Primus,E., Pyl,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojibokan,I., Rolfe,R., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shooshitari,N., Sisson,I.,
Soderren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Statak,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Umanil,K., Vasquez,L., Vera-V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
TITLE
JOURNAL
Unpublished
2 (bases 1 to 187332)
Worley,K.C.
Direct Submission
Submitted (12-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 187332)
Worley,K.C.
Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 12, 2002 this sequence version replaced gi:18847066.
COMMENT
--- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
--- Project Information
Center project name: GRIP
Center clone name: CH230-41K9
--- Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap: Version 0.990329
Consensus quality: 149700 bases at least Q40
Consensus quality: 156256 bases at least Q30
Consensus quality: 162284 bases at least Q20

```

*-----*
* NOTE: Estimated insert size may differ from sequence length.
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a working draft sequence. It currently
* consists of 56 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1001: contig of 1001 bp in length
* 1002 1101: gap of unknown length
* 1102 2590: contig of 1489 bp in length
* 2591 2690: gap of unknown length
* 2691 3916: contig of 1226 bp in length
* 3917 4016: gap of unknown length
* 4017 5147: contig of 1131 bp in length
* 5148 5247: gap of unknown length
* 5248 6451: contig of 1204 bp in length
* 6452 6551: gap of unknown length
* 6552 8012: contig of 1461 bp in length
* 8013 8112: gap of unknown length
* 8113 9240: contig of 1128 bp in length
* 9241 9340: gap of unknown length
* 9341 10703: contig of 1363 bp in length
* 10704 10803: gap of unknown length
* 10804 12877: contig of 2074 bp in length
* 12878 12977: gap of unknown length
* 12979 14298: contig of 1321 bp in length
* 14299 14398: gap of unknown length
* 14399 15844: contig of 1446 bp in length
* 15845 15944: gap of unknown length
* 15945 17587: contig of 1643 bp in length
* 17588 17687: gap of unknown length
* 17688 18733: contig of 1046 bp in length
* 18734 18833: gap of unknown length
* 18834 20948: contig of 2115 bp in length
* 20949 21048: gap of unknown length
* 21049 23317: contig of 2269 bp in length
* 23318 23417: gap of unknown length
* 23418 25429: contig of 2012 bp in length
* 25430 25529: gap of unknown length
* 25530 27057: contig of 1528 bp in length
* 27058 27157: gap of unknown length
* 27158 28820: contig of 1663 bp in length
* 28821 28920: gap of unknown length
* 28921 30258: contig of 1338 bp in length
* 30259 30358: gap of unknown length
* 30359 32574: contig of 2216 bp in length
* 32575 32674: gap of unknown length
* 32675 34314: contig of 1640 bp in length
* 34315 34414: gap of unknown length
* 34415 36024: contig of 1610 bp in length
* 36025 36124: gap of unknown length
* 36125 38547: contig of 2423 bp in length
* 38548 38647: gap of unknown length
* 38648 41952: contig of 3305 bp in length
* 41953 42052: gap of unknown length
* 42053 44168: contig of 2116 bp in length
* 44169 44268: gap of unknown length
* 44269 47028: contig of 2760 bp in length
* 47029 47128: gap of unknown length
* 47129 49048: contig of 1920 bp in length
* 49049 49148: gap of unknown length
* 49149 50748: contig of 1600 bp in length
* 50749 50848: gap of unknown length
* 50849 53484: contig of 2636 bp in length
* 53485 53584: gap of unknown length
* 53585 55862: contig of 2278 bp in length
* 55863 57817: gap of unknown length
* 57818 57917: gap of unknown length

```

```

* 57918 60973: contig of 3056 bp in length
* 60974 61073: gap of unknown length
* 61074 64269: contig of 3196 bp in length
* 64270 64369: gap of unknown length
* 64370 67983: contig of 3614 bp in length
* 67984 68083: gap of unknown length
* 68084 71383: contig of 3300 bp in length
* 71384 71483: gap of unknown length
* 71484 75737: contig of 4254 bp in length
* 75738 75837: gap of unknown length
* 75838 79520: contig of 3683 bp in length
* 79521 79620: gap of unknown length
* 79621 83422: contig of 3801 bp in length
* 83423 83522: gap of unknown length
* 83523 87515: contig of 3993 bp in length
* 87516 87615: gap of unknown length
* 87616 92111: contig of 4497 bp in length
* 92112 92211: gap of unknown length
* 92212 96534: contig of 4323 bp in length
* 96535 96634: gap of unknown length
* 96635 101680: contig of 5046 bp in length
* 101681 101780: gap of unknown length
* 101781 107239: contig of 5459 bp in length
* 107240 107339: gap of unknown length
* 107340 113255: contig of 5916 bp in length
* 113256 113355: gap of unknown length
* 113356 118197: contig of 4841 bp in length
* 118198 118296: gap of unknown length
* 118297 121989: contig of 3693 bp in length
* 121990 122089: gap of unknown length
* 122090 126511: contig of 4562 bp in length
* 126512 126751: gap of unknown length
* 126752 129765: contig of 3014 bp in length
* 129766 129865: gap of unknown length
* 129866 137350: contig of 7485 bp in length
* 137351 137450: gap of unknown length
* 137451 142649: contig of 5199 bp in length
* 142650 142749: gap of unknown length
* 142750 148513: contig of 5764 bp in length
* 148514 148613: gap of unknown length
* 148614 155533: contig of 6920 bp in length
* 155534 155633: gap of unknown length
* 155634 162354: contig of 6721 bp in length
* 162355 162454: gap of unknown length
* 162455 170099: contig of 7645 bp in length

Query Match      48.1%; Score 647.8; DB 2; Length 187332;
Best Local Similarity 68.6%; Pred. No. 2.3e-178;
Matches 1129; Conservative 0; Mismatches 212; Indels 304; Gaps 5;

```


VPLRCNTAPMMLNGSHPSSESRIVRTACAHMSDOCCNRSTELQVACRGGFTYIN
 LTAPECNLAACITDPSVSEGTRECEKRIEDICISDNGMRCKCKODSNITDVSLETRL
 ECGANDIKMSLRKCOLDSIFGNVFNVLNDROCSGSESEDERDMSIVTPARNPGCGT
 VLRRNEHATVSNITLYLANAIIIDIIIRNMFESYPLDMKVSLSLOPMSALNIS
 LGGTGKFTVYRMALFOSPTVTOPOGVSMLSTEAFLYVGTMLDGDLSRFLVLMNCS
 APTSSNSTDPVKYFIIDSCPRTEDTIOVTENGESSOARSVOMFRFAGNYDLYLH
 CEVYLCDSSTSECKPNCSTGSTRFRSGNFIIDOTRVNLGPIITRQGVQASVSASSNRL
 LSIWLLFISATLIFMYQ"

BASE COUNT 530 a 576 c 603 g 520 t
 ORIGIN
 Query Match 39.1%; Score 526.2; DB 10; Length 2229;
 Best Local Similarity 99.4%; Pred. No. 7.9e-143;
 Matches 528; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 816 GCAGAGCGGTGTTCTGATGCCAACACACCCACCTGCACGGTGATGTGTGTCACA 875
 DB 122 GCGAGACGGTCTTCGATGACCAACAACGCCACCTGCACGGTGATGTGTGTCACA 181
 QY 876 ACGTCTCTCTGCGACAGCGGCTTCACTGCTGATGGGCTGCTGTGTGAGACATGATGAG 935
 DB 182 ACGTCTCTCTGCGACAGCGGCTTCACTGCTGATGGGCTGCTGTGTGAGACATGATGAG 241
 QY 936 TGTGCTACCCCATGACTCACAACACTGCTCCACAGACAGCTGTGTGAACACCCCGGGCTCG 995
 DB 242 TGTGCTACCCCATGACTCACAACACTGCTCCACAGACAGCTGTGTGAACACCCCGGGCTCG 301
 QY 996 TTTAAGTGTCTCTGTCAGAGATGTTTTGCTGTGACGCTGAGCTGAGCTGACATGATGTG 1055
 DB 302 TTTAAGTGTCTCTGTCAGAGATGTTTTGCTGTGACGCTGAGCTGAGCTGACATGATGTG 361
 QY 1056 GATGAGTGTCTGACAGAGGGGCTCAGTAACGTGATGCCCTGGCCACCTGTGTCAACACA 1115
 DB 362 GATGAGTGTCTGACAGAGGGGCTCAGTAACGTGATGCCCTGGCCACCTGTGTCAACACA 421
 QY 1116 GAAGCGACTACTTGTGCTGTGTGCTGCTGCTTACAGGGGATGTTGGTACTGTGAG 1175
 DB 422 GAAGCGACTACTTGTGCTGTGTGCTGCTGCTTACAGGGGATGTTGGTACTGTGAG 481
 QY 1176 TGTGCTCCAGGCTCTGTGAGACGAGACTGACTGCTTGGCCAGGGCCGGATGGAAG 1235
 DB 482 TGTGCTCCAGGCTCTGTGAGACGAGACTGACTGCTTGGCCAGGGCCGGATGGAAG 541
 QY 1236 CTGCTGTGTCAAGACCCCTGCAATACATATGAGACCCCTGACAGTACTGTGGGACACACA 1295
 DB 542 CTGCTGTGTCAAGACCCCTGCAATACATATGAGACCCCTGACAGTACTGTGGGACACACA 601
 QY 1296 GAGTATGTTGGTGGCTACTCTCTGTGACGCGGGTCTGCACGGCTGTGACCGG 1346
 DB 602 GAGTATGTTGGTGGCTACTCTCTGTGACGCGGGTCTGCACGGCTGTGACCGG 652

RESULT 5
 MOSTAMM 2343 bp mRNA linear ROD 02-AUG-1995

LOCUS MUSTAMM 2343 bp mRNA linear ROD 02-AUG-1995
 DEFINITION Mus musculus uromodulin mRNA, complete cds.

ACCESSION L33406
 VERSION L33406.1 GI:927202

KEYWORDS Tamm-Horsfall; uromodulin.
 SOURCE Mus musculus cDNA to mRNA.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 2343)
 Prasadani, K., Bates, D., Badgett, A., Dell, M., Sukhame, V., Yu, H. and

Kuma, S.
 Nucleotide sequence and peptide motifs of mouse uromodulin

TITLE (Tamm-Horsfall protein)-the most abundant protein in mammalian

JOURNAL Biochem. Biophys. Acta 1260 (3), 328-332 (1995)

FEATURES Location/Qualifiers

1..2343
 /organism="Mus musculus"

mRNA
 5'UTR
 CDS
 /db_xref="taxon:10090"
 1..2343
 1..129
 130..2058

/standard_name="Tamm-Horsfall protein"
 /product="uromodulin"
 /protein_id="AA073896.1"
 /db_xref="GI:927203"

/translation="MGIPLTMLVMVTSNFTLAGASNSTEARCSECHNNATCTVD
 GVTYTSQGFPGFDGLVCEDEMDKATPWNTHNCSNVCNPFSGKSCODGRLTPG
 LGCTIDECSEOGISNCHALATVNTBEDGLVCEVCFKFTGDMYCCSPSSCEPRGLD
 LPQDPGLKIVCODPCNTPEYETLEFYWRSTSEYGVXCDAGDHGWYRPTGGGVMAETC
 VPIACNTAPMMLNGSHPSSESRIVRTACAHMSDOCCNRSTELQVACRGGFTYIN
 LTAPECNLAACITDPSVSEGTRECEKRIEDICISDNGMRCKCKODSNITDVSLETRL
 ECGANDIKMSLRKCOLDSIFGNVFNVLNDROCSGSESEDERDMSIVTPARNPGCGT
 VLRRNEHATVSNITLYLANAIIIDIIIRNMFESYPLDMKVSLSLOPMSALNIS
 LGGTGKFTVYRMALFOSPTVTOPOGVSMLSTEAFLYVGTMLDGDLSRFLVLMNCS
 APTSSNSTDPVKYFIIDSCPRTEDTIOVTENGESSOARSVOMFRFAGNYDLYLH
 CEVYLCDSSTSECKPNCSTGSTRFRSGNFIIDOTRVNLGPIITRQGVQASVSASSNRL
 LSIWLLFISATLIFMYQ"

BASE COUNT 555 a 604 c 631 g 553 t
 ORIGIN

Query Match 38.1%; Score 513.4; DB 10; Length 2343;
 Best Local Similarity 97.9%; Pred. No. 4.4e-139;
 Matches 520; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 816 GCAGAGCGGTGTTCTGATGCCAACACACCCACCTGCACGGTGATGTGTGTCACA 875
 DB 214 GCGAGAGGTGTTCTGAATGCCAACACACCCACCTGCACGGTGATGTGTGTCACA 273
 QY 876 ACGTCTCTCTGCGACAGCGGCTTCACTGCTGATGGGCTGCTGTGTGAGACATGATGAG 935
 DB 274 ACGTCTCTCTGCGACAGCGGCTTCACTGCTGATGGGCTGCTGTGTGAGACATGATGAG 933
 QY 936 TGTGCTACCCCATGACTCACAACACTGCTCCACAGACAGCTGTGTGAACACCCCGGGCTCG 995
 DB 334 TGTGCTACCCCATGACTCACAACACTGCTCCACAGACAGCTGTGTGAACACCCCGGGCTCG 993
 QY 996 TTTAAGTGTCTCTGTCAGAGATGTTTTGCTGTGACGCTGAGCTGAGCTGACATGATGTG 1055
 DB 394 TTTAAGTGTCTCTGTCAGAGATGTTTTGCTGTGACGCTGAGCTGAGCTGACATGATGTG 453
 QY 1056 GATGAGTGTCTGACAGAGGGGCTCAGTAACGTGATGCCCTGGCCACCTGTGTCAACACA 1115
 DB 454 GATGAGTGTCTGACAGAGGGGCTCAGTAACGTGATGCCCTGGCCACCTGTGTCAACACA 513
 QY 1116 GAAGCGACTACTTGTGCTGTGTGCTGCTGCTTACAGGGGATGTTGGTACTGTGAG 1175
 DB 514 GAAGCGACTACTTGTGCTGTGTGCTGCTGCTTACAGGGGATGTTGGTACTGTGAG 573
 QY 1176 TGTGCTCCAGGCTCTGTGAGACGAGACTGACTGCTTGGCCAGGGCCGGATGGAAG 1235
 DB 574 TGTGCTCCAGGCTCTGTGAGACGAGACTGACTGCTTGGCCAGGGCCGGATGGAAG 633
 QY 1236 CTGCTGTGTCAAGACCCCTGCAATACATATGAGACCCCTGACAGTACTGTGGGACACACA 1295
 DB 634 CTGCTGTGTCAAGACCCCTGCAATACATATGAGACCCCTGACAGTACTGTGGGACACACA 693
 QY 1296 GAGTATGTTGGTGGCTACTCTCTGTGACGCGGGTCTGCACGGCTGTGACCGG 1346
 DB 694 GAGTATGTTGGTGGCTACTCTCTGTGACGCGGGTCTGCACGGCTGTGACCGG 744

RESULT 6

LOCUS S75960 2211 bp mRNA linear ROD 27-JUL-1995

DEFINITION Tamm-Horsfall protein [rats, kidney, mRNA, 2211 nt].

ACCESSION S75960
 VERSION S75960.1 GI:912816

KEYWORDS

Wed Feb 19 14:09:14 2003

us-09-605-042a-1_copy-8000_9345.rge

SOURCE Rattus sp. kidney.
ORGANISM Rattus sp.; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mammalia; Rattus.

REFERENCE 1 (bases 1 to 2211)
AUTHORS Yu, H., Papa, F. and Sukhatme, V. P.
TITLE Bovine and rodent tamm-horsfall protein (THP) genes: cloning, structural analysis, and promoter identification
JOURNAL Gene Expr. 4 (1-2), 63-75 (1994)
MEDLINE 95143938
PUBMED
REMARK Genbank staff at the National Library of Medicine created this entry [NCBI g1bbsq 161879] from the original journal article. This sequence comes from Fig. 2B.

FEATURES
source location/Qualifiers
1..2211
/organism="Rattus sp."
/db_xref="taxon:10118"
1..2211
/gene="Tamm-Horsfall protein, THP"
16..1950
/gene="Tamm-Horsfall protein, THP"
/note="this sequence comes from Fig. 2B; THP"
/codon_start=1
/product="Tamm-Horsfall protein"
/protein_id="FAB3313.1"
/db_xref="GI:912817"
/translation="MSGQLSLTWLLVWVTPWTFVAGANDSPARSGSECHDNATCV
LDGVVTTSCQAGFTGDLGCEIDIECATPTWTHNCSNCTMNTIGSESCDGRFLT
PGLGCIQVNECTEBOGSLNCHSLATVETBSYSCVCPKGTGKQDHWGKRFYGGGVRAE
DCLPOGSGKLVCDPCNVEETLTYRSTDYAGVSCDSMDHGMSTETOVKACGGFYV
TCVPLRCNTAAPMLNLSHSPSSREGIVSRACAMSDHCLMSTETOVKACGGFYV
YLTPEPCNTLAYCTDPSVSEGTCEGVEDCVSDNGRMRCCOKODPNVNDVSLLEH
RLECEANEIKISLSKQLOSIGFPMKVFYLNDRCCSGFERGERDMSLYTPARAGPC
GVLRNRETHATYSNTLYLASELIIIDINIRINECSYPLDMKVSLSKTSLOPVASLN
ISLGTGKFTYOMALFONPTYTOPYOGSPWLSLEAFYVGMIDGDLSTREVLMTN
CYAPSSNSIDPYKFIIDRCPTEDDTTIOVENESSQARFSIOMREFAGNSDLY
CLATPSSNCTDPCPCGTCGTRYSGNFIIDQTRVLNLSGPIITGVOVASVKAASNL
LHCEVILCDTWSSECKPCGTCGTRYSGNFIIDQTRVLNLSGPIITGVOVASVKAASNL
GFLSIWLLFLSATIDTLNVH"

BASE COUNT 510 a 575 c 597 g 529 t

ORIGIN
Query Match 32.0%; Score 430.2; DB 10; Length 2211;
Best Local Similarity 88.1%; Pred. No. 9,6e-115; Indels 0; Gaps 0;
Matches 468; Conservative 0; Mismatches 63;

Y 816 GCAGAGCGGTGTTCTGAATGCCAACACCAACGCGCTGACGCGTGGATGTTGGTGCACA 875
106 GCAGAGAGGTGTTCTGAATGCCAACACCAACGCGCTGTTGGATGGGGTGCAC 935
Y 876 ACAGTCTCTCCAGACCGGCTTCACTGGTGGATGGGCTGTGTGAGACATGGATGAG 225
166 ACATGCTCTCCAGACGAGGCTTCACTGGATGGGCTGTGTGAGACATGGATGAG 225
Y 936 TGTGCTACCCAGTCACTCAACTGCTCCACAGCAGCTGTGTAACACCCGGGCTGC 285
226 TGTGCCACCCGCTGACACCAACTGCTCCACAGCAGCTGTGTAACACCCGGGCTGC 285
Y 996 TTTAAGTCTCTCTGAGAGATGTTTCTGTGACGCTGAGCTGAGATGAGATGATGTG 1055
286 TACGAGTGTCTCTGAGAGATGTTTCTGTGACGCTGAGCTGAGATGATGATGTG 345
Y 1056 GATGAGTGTCTGAGAGAGGCTTCACTGATGATGATGATGATGATGATGATGATGATG 1115
346 AATGAGTGTCTGAGAGAGGCTTCACTGATGATGATGATGATGATGATGATGATGATG 1175
Y 1116 GAAGGCACTACTGTGTGTGTGTCCGAGGCTTTACAGAGGATGTTGTACTGTGAG 1235
406 GAAGGCACTACTGTGTGTGTGTCCGAGGCTTTACAGAGGATGTTGTACTGTGAG 1235
Y 1176 TGTGCTCCAGGCTCTCTGAGAGATGATGATGATGATGATGATGATGATGATGATG 1235

Db 466 TGTCCCTCGGCTCTGTGAGACCGAGGCTGACTGCTTCCAGGCTCCAGGTCCAGTGAAG 525
Y 1236 CTGATGTGTCAAGACCCCTGCATATATAGACCCCTGACAGTACTGGCCACACACA 1295
526 CTGATGTGTCAAGACCCCTGCATATATAGACCCCTGACAGTACTGGCCACACACA 585
Y 1296 GAGTATGATGTGGTCTACTCTGTGACCGGGCTGACAGCGGCTGTACCGG 1346
586 GACTATGATGTGGTCTACTCTGTGACCGGGCTGACAGCGGCTGTACCGG 636

RESULT 7
RATUROMOD 2227 bp mRNA linear ROD 27-Apr-1993
LOCUS Rat Tamm-Horsfall protein mRNA, complete cds.
DEFINITION M63510
ACCESSION M63510.1 GI:207620
VERSION M63510.1 GI:207620
KEYWORDS Rat, cDNA to mRNA.
SOURCE Rattus norvegicus.
ORGANISM Rattus norvegicus; Chordata; Vertebrata; Euteleostomi; Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 2227)
AUTHORS Fukunaka, S., Freedman, S. D., Yu, H., Sukhatme, V. P. and Scheele, G. A.
TITLE GP-2/THP gene family encodes self-binding proteins in apical secretory glycosylphosphatidylinositol-anchored proteins in pancreatic compartments of pancreas and kidney
Proc. Natl. Acad. Sci. U.S.A. 89 (4), 1189-1193 (1992)

JOURNAL 92159014
MEDLINE 92159014
PUBMED 1515155

FEATURES
source location/Qualifiers
1..2227
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
1..2227
/db_xref="GI:207621"
/product="unimodulin"
/codon_start=1
/protein_id="AAA42319.1"
/translation="MSGQLSLTWLLVWVTPWTFVAGANDSPARSGSECHDNATCV
LDGVVTTSCQAGFTGDLGCEIDIECATPTWTHNCSNCTMNTIGSESCDGRFLT
PGLGCIQVNECTEBOGSLNCHSLATVETBSYSCVCPKGTGKQDHWGKRFYGGGVRAE
DCLPOGSGKLVCDPCNVEETLTYRSTDYAGVSCDSMDHGMSTETOVKACGGFYV
TCVPLRCNTAAPMLNLSHSPSSREGIVSRACAMSDHCLMSTETOVKACGGFYV
YLTPEPCNTLAYCTDPSVSEGTCEGVEDCVSDNGRMRCCOKODPNVNDVSLLEH
RLECEANEIKISLSKQLOSIGFPMKVFYLNDRCCSGFERGERDMSLYTPARAGPC
GVLRNRETHATYSNTLYLASELIIIDINIRINECSYPLDMKVSLSKTSLOPVASLN
ISLGTGKFTYOMALFONPTYTOPYOGSPWLSLEAFYVGMIDGDLSTREVLMTN
CYAPSSNSIDPYKFIIDRCPTEDDTTIOVENESSQARFSIOMREFAGNSDLY
CLATPSSNCTDPCPCGTCGTRYSGNFIIDQTRVLNLSGPIITGVOVASVKAASNL
LHCEVILCDTWSSECKPCGTCGTRYSGNFIIDQTRVLNLSGPIITGVOVASVKAASNL
GFLSIWLLFLSATIDTLNVH"

BASE COUNT 514 a 578 c 602 g 533 t

ORIGIN
Query Match 32.0%; Score 430.2; DB 10; Length 2227;
Best Local Similarity 88.1%; Pred. No. 9,6e-115; Indels 0; Gaps 0;
Matches 468; Conservative 0; Mismatches 63;

Y 816 GCAGAGCGGTGTTCTGAATGCCAACACCAACGCGCTGACGCGTGGATGTTGGTGCACA 875
115 GCAGAGAGGTGTTCTGAATGCCAACACCAACGCGCTGTTGGATGGGGTGCAC 935
Y 876 ACAGTCTCTCCAGACCGGCTTCACTGGTGGATGGGCTGTGTGAGACATGGATGAG 234
175 ACATGCTCTCCAGACGAGGCTTCACTGGATGGGCTGTGTGAGACATGGATGAG 234
Y 936 TGTGCTACCCAGTCACTCAACTGCTCCACAGCAGCTGTGTAACACCCGGGCTGC 294
235 TGTGCCACCCGCTGACACCAACTGCTCCACAGCAGCTGTGTAACACCACTGATG 1055
Y 996 TTTAAGTCTCTCTGAGAGATGTTTCTGTGACGCTGAGCTGAGCTGAGCTGATGTG 1055

Db 295 TAGCAGTGTCTCTGTCAGAGATGGCTTTGCTGTGAGCCCTGGGGCTGGGCTGCATTTGATGTG 354
Qy 1056 GATAGTGTCTAGAGACAGAGGGCTCACTAATGTGATCCCTGGCCACCTGTGTCAACACA 1115
Db 355 AATGATGTGACAGAGACAGAGGGCTCACTAATGTGATCCCTGGCCACCTGTGTCAACACG 414
Qy 1116 GAAGCGCAGCTACTGTGGTGTGTCGCCGAGGCTTACAGAGGGATGGTGTGTACTGTGAG 1175
Db 415 GAAGCGCAGCTACTGTGGTGTGTCGCCGAGGCTTACAGAGGGATGGTGTGTACTGTGAG 474
Qy 1176 TGTCTCCCGAGGCTCTGTGAGCCAGAGACTGTGCTTGGCCAGGCGCCGATGAGAAG 1235
Db 475 TGTCTCCCGAGGCTCTGTGAGCCAGAGGCTGTGCTTGGCCAGGCGCCGATGAGAAG 534
Qy 1236 CTGTGTGTCAAGACCCCTGCATATATGAGACCCCTGACTAGTACTGGCCAGACACA 1255
Db 535 CTGTGTGTCAAGACCCCTGCATATATGAGACCCCTGACTAGTACTGGCCAGACACA 594
Qy 1296 GAGTATGTGTGGGCTACTCTGTGAGCGGGGCTGTGACGGGTGTACCGG 1346
Db 595 GAGTATGTGTGGGCTACTCTGTGAGCTACATATGACAGCGGTGTACCGG 645

RESULT 8
AC106796/c 168887 bp DNA linear PRI 06-AUG-2002
LOCUS Homo sapiens chromosome 16 clone RP11-429K17, complete sequence.
DEFINITION AC106796
AC106796.2 GI:22122880
VERSION
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 168887)
TITLE DOE Joint Genome Institute.
JOURNAL Sequencing of Human Chromosome 16
AUTHORS Unpublished
2 (bases 1 to 168887)
TITLE DOE Joint Genome Institute.
JOURNAL Direct Submission
AUTHORS Submitted (12-JAN-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 168887)
TITLE DOE Joint Genome Institute.
JOURNAL Direct Submission
AUTHORS Submitted (06-AUG-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Aug 6, 2002 this sequence version replaced gi:18139346.
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
Genbank flat file format but are available as part
of this entry's ASN.1 file.

Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
Genbank flat file format but are available as part
of this entry's ASN.1 file.

FEATURES
source Location/Qualifiers
1..168887
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 46572 a 37397 c 37923 g 46995 t
ORIGIN
Query Match 23.1%; Score 311.2; DB 9; Length 168887;
Best Local Similarity 74.2%; Pred. No. 7.1e-80;
Matches 423; Conservative 0; Mismatches 138; Indels 9; Gaps 2;
Qy 779 CCAGCTGCAAGGATCTGAAGCTGGGCTCTTCTGTCCGACAGAGGCTTCTGAATGCCA 838
Db 116770 CCAGCTGCCAACCCTGAAGCTGGGCTTCTCTCCACAGAGATGCTCTGAATGTCA 116711
Qy 839 CAACCAAGCCACTGCACGGTGTGAGTGTGTGTACAAAGTCTCTCCGACAGCGGCTT 898
Db 116710 CAGCAATGTCCACTGCACGGATGAGGAGGAGCCGTTTACAGAGCTGTCTCAGAGGCTT 116651
Qy 899 CACTGTGATGAGGCTGTGTGTGTGAGACATGATGATGTCTACCCCATGACTACAA 958
Db 116650 CACCGCGCATGGCTGTGACTGCGGTGAGACCTGATGATGAGGCGCATCTCGAGCTCACAA 116591
Qy 959 CTGCT---CCACAGCAGCTGTGTGAACACCCCGGCTCGTTTAAGTCTCTGTACGA 1015
Db 116590 CTGCTCCGCAACAGCAGCTGTGCTTAAACAGCCAGGCTCTTCTCTGCTGCTGCCCA 116531
Qy 1016 TGGTTTTCGTGAGCGCTGAGCTGAGCTGCATGATGTGATGTGATGAGTGCAGACAGG 1075
Db 116530 AGGCTTCGCGCTGTGCGCGGCTGTGCTGTGACAGACCTGTGATGATGCTGTGAGCTCG 116471
Qy 1076 GCTCACTACTGTGATGATGCTGTGCGCACCTGTGTGACACAAAGCGACTACTGTGCGT 1135
Db 116470 GCTTACCACTGTGACAGCCCTGTGCGCATGTGATGTGAGGAGCAGCTACTGTGCGT 116411
Qy 1136 GTGTCCCGAAGGCTTTTACAGAGGATGTGTGTACTGTAGTCTCTCCAGGCTCTGTGA 1195
Db 116410 ATGCCCGGCGGCTTACCGGGGGGTGATGTGACTGTGATGCTCCCGGCTCTCGCG 116351
Qy 1196 GCCAGAGCTGAGCTGTGTCGCCCAAGGCGCGGATGGAAGAGTGGTGTCAAGACCCCTG 1255
Db 116350 GCCGGGTTGGACTGTGCTGCTCCAGGCGC-----GACCGCTCGTGTGCGGATCCG 116297
Qy 1256 CAATACATATGAGACCCCTGATGATGATGCGCAGCAGAGATGATGTGTGGCTACTC 1315
Db 116296 CCAGCGCAGCCGACCTGTGACAGTACTGTGCGCAGCAGCAGATGCGGGAGGCTACCG 116237
Qy 1315 CTGTGACGCGGGTCTGTGACGCGCTGTACCG 1345
Db 116236 CTGCGACAGGACCTGCGCGCTGTGTACCG 116207

RESULT 9
HUMMODA
LOCUS HUMMODA 2290 bp mRNA linear PRI 03-AUG-1993
DEFINITION Human uromodulin (Tamm-Horsfall glycoprotein) mRNA, complete cds.
ACCESSION M17778
VERSION M17778.1 GI:340165
KEYWORDS Tamm-Horsfall glycoprotein; glycoprotein; uromodulin.
SOURCE Human kidney, cdna to mRNA, clone 6.7.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2290)
AUTHORS Hession, C., Decker, J.M., Sherblom, A.P., Kumar, S., Yue, C.C.,
Mattaliano, R.J., Tizard, R., Kawashima, E., Schneissner, U.,
Heletky, S., Chow, E.P., Burne, C.A., Shaw, A. and Muchmore, A.V.
Uromodulin (Tamm-Horsfall glycoprotein): a renal ligand for
lymphokines
JOURNAL Science 237 (4821), 1479-1484 (1987)
MEDLINE 87319675
PUBMED 3498215
FEATURES
source Location/Qualifiers
1..2290
/organism="Homo sapiens"

mrna
CDS
/db_xref="taxon:9606"
1..2290
/product="uromodulin mRNA"
106..2028
/note="uromodulin"
/codon_start=1
/protein_id="AAA36799.1"
/db_xref="GI:340166"
/translation="MGPSLTWMLVYVVAWMTTATATDTSSEARMCSECHSNATCTED
EAVTTCQEGFTGDLTCDVDECAIPAHNCSANSVNTPGSFSCVCEGRSLP
GLGCTDVEACAPGLSHCHALATCVNVVSGSYLCVCPAGYGDGWHCEGSPGCGPLD
CVPEDALVACAPCAHRTLDYXRSTREYEGTACTDLDGMYRFPVGGARMEATV
PVLKNTAPAMWLNTPSSDEGIVSRKACAHWSGCCLDWASQVACAGGYVYNL
TAPPCCHALCTDSSVEGCECSIDEDCKSNRNHCCKODFNITDISLEHRL
CGANMKVSLGCKQKSLGFDKRYMILSDRSGNDNRYMVSVPARDPGCV
LFRNTHATYSLNLYLADELITDNLKINPACYPIDMKVSLKALQPNVSLNIRV
GGTGMFTVMALFQPSYTOPYGSVYLTSTEARLYGTMLDGDLSHFALLMTYCA
TPSSNATDLKFTIIDRCFHTDSTIQVVENSSGKRSVQMFPRFAGNTDLYLDC
EYICDPMNEKCKPTCGTRFRSGSYIDOSRVNLGPIITRKGVQATVSRNLSGLLK
VWPLLSATLITLTFQ"

BASE COUNT 479 a 675 c 660 g 476 t
About 25 bp downstream of 5' end of mRNA.

ORIGIN
Query Match 21.1%; Score 283.8; DB 9; Length 2290;
Best Local Similarity 73.5%; Pred. No. 6.6e-72;
Matches 392; Conservative 0; Mismatches 132; Indels 9; Gaps 2;

816 GCAGGACGCTGTTCTGAATGCACAAACACGACCTGCACGCTGATGCTGTCACA 875
190 GCAAGATGCTGCTGATGATGACAGCAATGCCACTGCACGAGGATGAGCCGTTACG 249
876 ACGTCTCTCCAGACCGGCTTCACTGCTGATGGCTGTGTGAGACATGATGAG 935
250 ACGTCACTCTGACGAGGAGGCTTCAACCGGATGGCTTCACTGCTGAGACCTGGATGAG 309
936 TGTGCTACCCCATGAGACTCAACTGCT---CCACAGCAGCTGTGTGAACACCCCGGGC 992
310 TCGCCATCTCCGAGCTCAACTGCTCCGCCAAGCAGCTGCTGAACACGCGCAGG 369
993 TCGTTAAGTCTCTGTCAGAGTGTGTTGCTGTCAGCTGAGCTGAGTGCATGAT 1052
370 TCTCTTCTCTGCTGCTGCGCCGAGGCTTCCGCTGCTGCGCGGCTGCTGCTGACACAG 429
1053 GTGATGAGTGTCTCAAGACGAGGCTCACTGATGCTGCTGCTGCTGCTGCTGCTGCTG 1112
430 GTGATGAGTGTCTCAAGACGAGGCTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 489
1113 ACAGAGGCGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1172
490 GTGCTGCGCAGCTACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 549
1173 GAGTGTCTCCAGAGCTCTCTGAGCCAGAGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1232
550 GAGTGTCTCCAGAGCTCTCTGAGCCAGAGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 603
1233 AAGCTGTGTGTCAAGACCCCTGCAATATATGAGACCTGACTGACTGAGTGGCGAGC 1292
604 GCGCTGCTGTGCGCGGAGTCCGTGTCAGCGCAGCCGACCTGGACAGTACTGGCGAGC 663
1293 ACAGAGTATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1345
664 ACCGAGTACGGGAGGCTTACGCTGCGACAGGACCTGCGCGCTGTGTACCG 716

RESULT 10
AX336081. 2353 bp DNA linear PAT 09-JAN-2002
LOCUS AX336081
DEFINITION Sequence 6590 from Patent WO0194629.
ACCESSION AX336081
VERSION AX336081.1 GI:18126800
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
BASE COUNT 493 a 692 c 673 g 495 t
Query Match 21.1%; Score 283.8; DB 6; Length 2353;
Best Local Similarity 73.5%; Pred. No. 6.6e-72;
Matches 392; Conservative 0; Mismatches 132; Indels 9; Gaps 2;

816 GCAGGACGCTGTTCTGAATGCACAAACACGACCTGCACGCTGATGCTGTCACA 875
253 GCAAGATGCTGCTGATGATGACAGCAATGCCACTGCACGAGGATGAGCCGTTACG 312
876 ACGTCTCTCCAGACCGGCTTCACTGCTGATGGCTGTGTGAGACATGATGAG 935
313 ACGTCACTCTGACGAGGAGGCTTCAACCGGATGGCTTCACTGCTGAGACCTGGATGAG 372
936 TGTGCTACCCCATGAGACTCAACTGCT---CCACAGCAGCTGTGTGAACACCCCGGGC 992
373 TCGCCATCTCCGAGCTCAACTGCTCCGCCAAGCAGCTGCTGCTGAACACGCGCAGG 432
993 TCGTTAAGTCTCTGTCAGAGTGTGTTGCTGTCAGCTGAGCTGAGCTGACTGATGAT 1052
433 TCTCTTCTCTGCTGCTGCGCCGAGGCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 492
1053 GTGATGAGTGTCTCAAGACGAGGCTCACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1112
493 GTGATGAGTGTCTCAAGACGAGGCTTCACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 552
1113 ACAGAGGCGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1172
553 GTGCTGCGCAGCTACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 612
1173 GAGTGTCTCCAGAGCTCTCTGAGCCAGAGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1232
613 GAGTGTCTCCAGAGCTCTCTGAGCCAGAGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 666
1233 AAGCTGTGTGTCAAGACCCCTGCAATATGAGACCTGACTGACTGAGTGGCGAGC 1292
667 GCGCTGCTGTGCGCGGAGTCCGTGTCAGCGCAGCCGACCTGGACAGTACTGGCGAGC 726
1293 ACAGAGTATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1345
727 ACCGAGTACGGGAGGCTTACGCTGCGACAGGACCTGCGCGCTGTGTACCG 779

RESULT 11
AX336346 2353 bp DNA linear PAT 09-JAN-2002
LOCUS AX336346
DEFINITION Sequence 6855 from Patent WO0194629.
ACCESSION AX336346
VERSION AX336346.1 GI:18127065
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
BASE COUNT 493 a 692 c 673 g 495 t
Query Match 21.1%; Score 283.8; DB 6; Length 2353;
Best Local Similarity 73.5%; Pred. No. 6.6e-72;
Matches 392; Conservative 0; Mismatches 132; Indels 9; Gaps 2;

816 GCAGGACGCTGTTCTGAATGCACAAACACGACCTGCACGCTGATGCTGTCACA 875
253 GCAAGATGCTGCTGATGATGACAGCAATGCCACTGCACGAGGATGAGCCGTTACG 312
876 ACGTCTCTCCAGACCGGCTTCACTGCTGATGGCTGTGTGAGACATGATGAG 935
313 ACGTCACTCTGACGAGGAGGCTTCAACCGGATGGCTTCACTGCTGAGACCTGGATGAG 372
936 TGTGCTACCCCATGAGACTCAACTGCT---CCACAGCAGCTGTGTGAACACCCCGGGC 992
373 TCGCCATCTCCGAGCTCAACTGCTCCGCCAAGCAGCTGCTGCTGAACACGCGCAGG 432
993 TCGTTAAGTCTCTGTCAGAGTGTGTTGCTGTCAGCTGAGCTGAGCTGACTGATGAT 1052
433 TCTCTTCTCTGCTGCTGCGCCGAGGCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 492
1053 GTGATGAGTGTCTCAAGACGAGGCTCACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1112
493 GTGATGAGTGTCTCAAGACGAGGCTTCACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 552
1113 ACAGAGGCGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1172
553 GTGCTGCGCAGCTACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 612
1173 GAGTGTCTCCAGAGCTCTCTGAGCCAGAGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1232
613 GAGTGTCTCCAGAGCTCTCTGAGCCAGAGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 666
1233 AAGCTGTGTGTCAAGACCCCTGCAATATGAGACCTGACTGACTGAGTGGCGAGC 1292
667 GCGCTGCTGTGCGCGGAGTCCGTGTCAGCGCAGCCGACCTGGACAGTACTGGCGAGC 726
1293 ACAGAGTATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1345
727 ACCGAGTACGGGAGGCTTACGCTGCGACAGGACCTGCGCGCTGTGTACCG 779

EAVTTCQEGFTGDLTCVDLDECAIPGAHNCANSNCVNTPGSESCVCEGFRISP

RESULT 15
LOCUS S75958 2340 bp mRNA linear MAM 27-JUL-1995
DEFINITION Tamm-Horsfall protein [cattle, kidney, mRNA, 2340 nt].
ACCESSION S75958
VERSION S75958.1 GI:912814
KEYWORDS
SOURCE Bos taurus kidney.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 2340)
Yu, H., Papa, F. and Sukhatme, V.P.
Bovine and rodent tamm-horsfall protein (THP) genes: cloning,
structural analysis, and promoter identification
Gene Expr. 4 (1-2), 63-75 (1994)
JOURNAL MEDLINE 95143938
PUBMED 7531049
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI 91bbsq 161877] from the original journal article.
This sequence comes from Fig. 2A.
FEATURES
source
1..2340
location/Qualifiers
/organism="Bos taurus"
/db_xref="taxon:9913"
1..2340
/gene="Tamm-Horsfall protein, THP"
118..2049
/gene="Tamm-Horsfall protein, THP"
/note="this sequence comes from Fig. 2A, THP"
/codon_start=1
/product="Tamm-Horsfall protein"
/protein_id="AAB3312.1"
/db_xref="GI:912815"
translation="MKCLEPSPFMMAAVTVSVIIPAAITPSSAKSCSECHSNATCT
VDGAATTCACOBEGTDECDLECYDLDECAVGLAHNCATKSCVATLSTVCPEPGLL
SSELCEDVDDECAPEGLSRCHALATCINGENISVCVPAGYLDGRHCECSPGSCGP
LDVAREGDALVCPCQVHRIIDEXWSTEGSYICDYSIGMYRFVQAGVRLPET
CVPVLCNTAAPMMLNGTHPSSDEGIVNRVACAHMSGDCCLMDAPIOVKACAGGYVY
NLTAPPECHLAICTDPSSVEGCECRVEDKSDNGEMHCCCKODPNVTDLSLEBR
LEGVDVIXISKCOLKSLGFEKXEMVYHDSCSGFTREKQDRDMSVYTPADPGCG
TVATRNETHATYSNVIYLADEIIRDLNTRINACSYPIIDMKVSLKTSLOPMVSAINI
SMGGTGTFTYMALEFQSPAYTQPYGSSVTLSTEARLYVGTMLDGDLSRFVILMNC
YAPSSNATYDPLKFIIDRCPRADSTIOVENESPOGRFSPQMFREFRNVLDVYL
HCEVYLCDTVNEKCRPTCEPTFRFRSGSIIDQTRVILNLPITTRKGQAAMSRAPSSLG
LLOVWMLPLLSATLTLMSP"

BASE COUNT 504 a 666 c 661 g 509 t
ORIGIN

Query Match 19.0%; Score 256.4; DB 4; Length 2340;
Best Local Similarity 70.1%; Pred. No. 6.9e-64;
Matches 376; Conservative 0; Mismatches 151; Indels 9; Gaps 2;

QY 813 TCOCAGAGAGCGTGTCTGAATGCCACAACACGCCACTGCACGGTGAATGTTGTC 872
DB 205 TCAGCAAAAACGCTCTGTAATGTACAGCAATGCCACTGTACGGTGAAGGGGGCTGCC 264
QY 873 ACAAGTGTCTCTCCAGACCGGCTTCACTGGTGAATGGGCTGGTGTGAGACATGGAT 932
DB 265 ACAGACCTGCGCGCTCCAGAGAGGGCTTCACTGGCGAGCGGCTCGAGTGTGATCTGGAC 324
QY 933 GAGTGTCTACCCCATGAGCTACAACTGCT--CCACAGCAGCTGTGTGAACACCCCG 989
DB 325 GAATGCCCGCTTCTGGGGGCGCACAACTGCTCCGCCACCAAGAGCTGCGTGAATAGCTG 384
QY 990 GGGCTGTTAAGTCTCTCTGATGATGTTTCTGCTACAGCCCTAGCTGAGCTGCACT 1049
DB 385 GGGCTTTACAGAGTGTCTGCTGAGTTTCTGCTGAGCTCGGAGCTCGGCTGGAG 444
QY 1050 GATGTGATGATGCTCTAGACAGAGGGCTCACTAATGTCATGCCCTGGCCACCTGTGTC 1109

DB 445 GATGTGACAGAGTGTGACAGAGCCAGGGCTCAGCCGCTGCCACCCCTGGCCACTTGATC 504
QY 1110 AACACAGAGCGGACTACTTGTGCTGTGTCCCGAGGGCTTACAGGGAGTGTGTAC 1169
DB 505 AATGGCAGAGGCACTACTCTGCTGTGTCTGTGCGGCTACTGGGAGACGGAAGCAC 564
QY 1170 TGTGAGTCTTCCCGAGGCTCTGTGAGCCAGACTGAGTGTGCTTGGCCCAAGGCCGAT 1229
DB 565 TGTGAGTCTTCCCGAGGCTCTGTGAGCCAGACTGAGTGTGCGGAGGCGC----- 618
QY 1230 GAAAGTGTGTGTCAAGACCCCTGCAATACATATGAGACCTGTGATCTGGCGC 1289
DB 619 GAGCGCTGTGTCTGCTGAGACCCCTGCGACAGGTGCACCCGATCTGAGCAATATCTGGCGC 678
QY 1290 AGCAGAGATGATGTGTGAGGCTACTCTGTGAGCGCGGCTGTGCACGCTGTACCG 1345
DB 679 AGCAGAGATGAGGCTCGGCTACTATCTGTATGTCAAGTCTGGCGGCTGTACCG 734

Search completed: February 17, 2003, 15:15:26
Job time : 4011.36 secs

Wed Feb 19 14:09:08 2003

us-09-605-042a-1-copy_4000_6000.rge

Gencore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 17, 2003, 08:20:43 ; Search time 4997.09 Seconds
(without alignments)
11653.717 Million cell updates/sec

US-09-605-042a-1-copy_4000_6000

Title: 2001
Sequence: 1 aagtaacctgttataaca.....tgtatataatacatataca 2001Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

4109280

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000Post-processing: Minimum Match 100%
Listing first 45 summaries

Database :

GenBank:*

1: gb_ba:*

2: gb_hg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pac:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vl:*

30: em_hg_hum:*

31: em_hg_inv:*

32: em_hg_other:*

33: em_hg_mus:*

34: em_hg_pln:*

35: em_hg_rtd:*

36: em_hg_mam:*

37: em_hg_vrt:*

38: em_sy:*

39: em_hggo_hum:*

40: em_hggo_mus:*

41: em_hggo_other:*

Pred No is the number of results predicted by chance to have a

score greater than or equal to the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1976.4	98.8	219386	2 AC125147	AC125147 Mus muscu
2	924.8	46.2	187332	2 AC110401	AC110401 Rattus no
3	617.6	30.9	165503	2 AC111518	AC111518 Rattus no
4	101.2	5.1	156954	2 AL831768	AL831768 Danio rer
5	97.4	4.9	168153	2 AL772289	AL772289 Danio rer
6	86.6	4.3	269081	2 AC068493	AC068493 Mus muscu
7	86.6	4.3	179047	2 AC123466	AC123466 Rattus no
8	86.6	4.3	7218	6 AC124397	AC124397 Rattus no
9	79.8	4.0	227271	2 AC122205	AC122205 Mus muscu
10	78.2	3.9	123116	2 AC100738	AC100738 Mus muscu
11	76.8	3.8	70420	2 AC121139	AC121139 Mus muscu
12	75.8	3.8	145883	2 AC128306	AC128306 Rattus no
13	75.6	3.8	172121	2 AL845338	AL845338 Mus muscu
14	75.6	3.7	12121	2 AC124397	AC124397 Mus muscu
15	74.8	3.7	184358	2 AL844571	AL844571 Mus muscu
16	74.6	3.7	179255	2 AC127923	AC127923 Rattus no
17	74.4	3.7	182994	2 AC093290	AC093290 Homo sapi
18	74.4	3.7	177639	2 AC121603	AC121603 Mus muscu
19	74.2	3.7	193690	2 AC121823	AC121823 Mus muscu
20	74.2	3.7	214793	2 AC118017	AC118017 Mus muscu
21	74.2	3.7	177412	2 AL844171	AL844171 Mus muscu
22	74	3.7	182848	2 AC073784	AC073784 Mus muscu
23	74	3.7	235302	2 AC004773	AC004773 Oryza sat
24	73.8	3.7	136073	2 AC121738	AC121738 Danio rer
25	73.8	3.7	163190	2 AC125498	AC125498 Danio rer
26	73.4	3.7	148115	2 AL833789	AL833789 Danio rer
27	73.4	3.7	172462	2 AL808141	AL808141 Danio rer
28	72.8	3.6	147212	2 AC131362	AC131362 Rattus no
29	72.6	3.6	113894	2 AC131362	AC131362 Rattus no
30	72.6	3.6	240871	2 AC122568	AC122568 Rattus no
31	72.6	3.6	181350	10 AL772214	AL772214 Mouse DNA
32	72.4	3.6	230097	2 AC126033	AC126033 Mus muscu
33	72.4	3.6	170562	2 AC096985	AC096985 Rattus no
34	72	3.6	164876	2 AL627167	AL627167 Danio rer
35	71.8	3.6	194454	2 AC127295	AC127295 Mus muscu
36	71.8	3.6	220778	2 AC124757	AC124757 Mus muscu
37	71.8	3.6	241985	2 AC124584	AC124584 Mus muscu
38	71.6	3.6	241985	2 AL844217	AL844217 Danio rer
39	71.6	3.6	237550	2 AL845325	AL845325 Mus muscu
40	71.2	3.5	203739	2 AC102635	AC102635 Mus muscu
41	70.8	3.5	158266	2 AC109704	AC109704 Rattus no
42	70.8	3.5	160122	2 AC097697	AC097697 Rattus no
43	70.6	3.5	58079	2 AC106363	AC106363 Rattus no
44	70.6	3.5	216774	2 AC095949	AC095949 Rattus no
45					

ALIGNMENTS

RESULT 1

AC125147

LOCUS AC125147

DEFINITION Mus musculus chromosome UNK clone RP24-350A15, WORKING DRAFT

SEQUENCE 8 unordered pieces.

ACCESSION AC125147

VERSION AC125147.2 GI:21539180

HTG: HTGS-PHASE1; HTGS-DRAFT.

KEYWORDS

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 219386)

AUTHORS McPherson J.D. and Waterston R.H.

TITLE The sequence of Mus musculus clone

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Unpublished
2 (bases 1 to 219386)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (20-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 219386)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (22-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
On Jun 22, 2002 this sequence version replaced gi:21490647.

```

-----
Center: Washington University Genome Sequencing Center
Center code: WDGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.edu
-----
Project Information
Center project name: M_BB0350A15
-----
Summary Statistics
-----

```

```

Sequencing vector: plasmid; 100%
Chemistry: Dye-Primer ET; % of reads
Chemistry: Dye-Terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 216121 bases at least Q40
Consensus quality: 216506 bases at least Q40
Consensus quality: 216719 bases at least Q20
Insert size: 11000; agarose-fp
Insert size: 217980; sum-of-contigs
Quality coverage: 14.92 in Q20 bases; agarose-fp
Quality coverage: 11.97 in Q20 bases; sum-of-contigs

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

*	1	105: contig of 105 bp in length
*	106	205: gap of unknown length
*	206	806: contig of 601 bp in length
*	807	906: gap of unknown length
*	907	5269: contig of 4363 bp in length
*	5270	5369: gap of unknown length
*	5370	16399: contig of 11024 bp in length
*	16394	16493: gap of unknown length
*	16494	34651: contig of 18158 bp in length
*	34652	34751: gap of unknown length
*	34752	80845: contig of 46094 bp in length
*	80846	80945: gap of unknown length
*	80946	138244: contig of 57297 bp in length
*	138245	138342: gap of unknown length
*	138243	219368: contig of 81044 bp in length
*	138343	

FEATURES	Location/Qualifiers
source	1. .219386

```
misc_feature 1..105 /note="assembly_name:Contig12"
misc_feature 206..806 /note="assembly_name:Contig14"
misc_feature 907..5269 /note="assembly_name:Contig19"
misc_feature 5370..16393 /note="assembly_name:Contig20"
misc_feature
```

```

misc_feature      /note="assembly_name:Contig21"
                  34752. 80845
misc_feature      /note="assembly_name:Contig22"
                  80946. 138242
misc_feature      /note="assembly_name:Contig23"
                  138343. 219386
misc_feature      /note="assembly_name:Contig24"
                  60953 a 47433 c 47470 g 62799 t 731 others
BASE COUNT
ORIGIN
Query Match      98.8%; Score 1976.4; DB 2; Length 219386;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1999; Conservative 0; Mismatches 1; Indels 2; Gaps 2

```

QY	1	AAAGTACTCTGTTATACATCCATATCAAAATGCGCCCTGTGTAGCAAGTTAAAG	60
Dd	122697	AACTAAGTCTGTTATTAACAAATCCATATCAAAATGCGCCCTGTGTAGCAAGTTAAAG	122756
QY	61	AAAAGATTATCCAGAGATCCAAAGTCTCTTCAAAACCAAGTGTACAGCAATTGTC	120
Dd	122757	AAAAGATTATTCAGAGATCCAAAGTCTCTTCAAAACCAAGTGTGTACAGCAATTGTC	122816
QY	121	TCAGAGTAAAGATTGCAATTTGGCAACATGCAATGCTTTAAAGGTGTGGG - GAATTTCAGT	179
Dd	122817	TCAGAGTAAAGATTGCAATTTGGCAACATGCAATGCTTTAAAGGTGTGGG - GAATTTCAGT	122876
QY	180	GGAGTTGGACGTCAGAAACACACACTGCTGTAATAATGGAGAGAAATAGATTATTCCTTTGA	239
Dd	122877	GGAGTTGGACGTCAGAAACACACACTGCTGTAATAATGGAGAGAAATAGATTATTCCTTTGA	122938
QY	240	GAATTTGGTCTCAAAAAAGTAGGTATCAAAATTACTTGTGTCTGTGAGATCAATTGGTT	299
Dd	122937	GAATTTGGTCTCAAAAAAGTAGGTATCAAAATTACTTGTGTCTGTGAGATCAATTGGTT	122996
QY	300	GTCTCTGATAGTTAGCTTTACATAGAGACAGAAATAGTGAAGAGAGAGAGAGACAT	359
Dd	122997	GTCTCTGATAGTTAGCTTTACATAGAGACAGAAATAGTGAAGAGAGAGAGAGACAT	123056
QY	360	TGGAGCACCCAGAGAGAGAGAGACCTTCTCTCTTAAATGAAATAGGTGTGCTTCATTCC	419
Dd	123057	TGGAGCACCCAGAGAGAGAGAGACCTTCTCTCTTAAATGAAATAGGTGTGCTTCATTCC	123118
QY	420	AAAGACAAGAGATTCAAGTGTGCGCCGGGAAGATGAGAGACCAATCCACAAAGAAATGGCA	479
Dd	123117	AAAGACAAGAGATTCAAGTGTGCGCCGGGAAGATGAGAGACCAATCCACAAAGAAATGGCA	123176
QY	480	GGAAGTCACTCTGTGTCATTAATGAGAGAGGGGCTCAAAAGTGGACCAAGAGATG	539
Dd	123177	GGAAGTCACTCTGTGTCATTAATGAGAGAGGGGCTCAAAAGTGGACCAAGAGATG	123238
QY	540	AGCAATAAAATGGTGGATGTGGATACACTGTAGAGATGGCTGGCTGTGGACCAAAATGT	599
Dd	123237	AGCAATAAAATGGTGGATGTGGATACACTGTAGAGATGGCTGGCTGTGGACCAAAATGT	123296
QY	600	GGGCAAGTGGCACTCCATTAACAAAGACAGCTGCTGTGTGGACATCTTAAATAAAG	659
Dd	123297	GGGCAAGTGGCACTCCATTAACAAAGACAGCTGCTGTGTGGACATCTTAAATAAAG	123356
QY	660	GCACATGGCATGCCATGGAAGGCTTAGGGAGTTGAGGGGAAAGGTATATAGATAGATGCA	719
Dd	123357	GCACATGGCATGCCATGGAAGGCTTAGGGAGTTGAGGGGAAAGGTATATAGATAGATGCA	123416
QY	720	AACTACAGAGAGAGCCAGGAAGAGACAGAGATAGAGAGGAGACAGTTTGCACAAAGCCTTGT	779
Dd	123417	AACTACAGAGAGAGCCAGGAAGAGACAGAGATAGAGAGGAGACAGTTTGCACAAAGCCTTGT	123476
QY	780	CCCTCCCAAGAGAGCTTCTCCCTTCTGTATATGCAATACACAGTAGCTACTGTGC	839
Dd	123477	CCCTCCCAAGAGAGCTTCTCCCTTCTGTATATGCAATACACAGTAGCTACTGTGC	123536
QY	840	AAATGCTGCACATATGATGATGATGAAACAGAGCCACTGTTGGGTACAGTCTTACAGC	899
Dd	123537	AAATGCTGCACATATGATGATGATGAAACAGAGCCACTGTTGGGTACAGTCTTACAGC	123596

QY	1980	TATGTATATATATACATACA	2001
Db	124676	TATGTATATATATACATACA	124697
RESULT 2			
AC110401			
LOCUS			
DEFINITION	AC110401	187332 bp	DNA linear HTG 13-JUL-2002
ACCESSION	AC110401		
VERSION	AC110401.3	GI:21738952	
KEYWORDS	HTG; HTGS PHASE1.		
SOURCE	Normax rat.		
ORGANISM	Rattus norvegicus		
	Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
REFERENCE	1 (bases 1 to 187332)		
AUTHORS	Munzy,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alstrooks,S.L., Amaralunge,H.C., Are,J.R., Ayale,M., Banks,T., Barbarta,J., Benton,J., Bimaga,K., Blankenburg,K., Bonin,D., Bouck,J., Bowle,S., Brieval,M., Brown,M., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chaver,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Day-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B., Homsli,F., Howard,S., Huber,J., Huliyil,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Kovach,J., Kovar,C., Kratochvic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisedge,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mashinney,E., Mcleod,M.P., Meador,M., Mei,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwenkwo,S., Oguh,M., Okwuonu,G., Ogunyemi,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojka,A., Rojupokan,I., Rolfe,M., Ruiz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shooshtrai,N., Sisson,T., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tanami,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,Y., Villalón,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooten,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.		
JOURNAL	Unpublished		
TITLE	Direct Submission		
AUTHORS	2 (bases 1 to 187332)		
JOURNAL	Worley K.C.		
TITLE	Direct Submission		
AUTHORS	Submitted (12-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA		
REFERENCE	3 (bases 1 to 187332)		
AUTHORS	Worley K.C.		
TITLE	Direct Submission		
JOURNAL	Submitted (13-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One		

OY	126	AATAAGATTGCACTTTGGCAACATGCATGTCTTATGTGGTGGA- AATTTAGTAGCAT	184
Db	176630	AGTAAGACTGCACTTTGACAACATGCAATGTCTTTGACAGTGTGGCTAATTTAGTGAAGA	176689
OY	185	TGGCAGCTGAGAAAAGCACACTGTGTAAAAAATGAGAGAAATGATATATCTTTGAGAAAT	244
Db	176690	TGGCACTTCAGAAAAGCAACATCGTGGTGAATTTGAGAGAGTGAATAACTCTTTGGGAAAT	176749
OY	245	TTGGTCTCAAAGTAAGGGATATCAAAATTTACTTGGTGTCTGTGAATTCATTTGGTGTCTC	304
Db	176750	TTGGTCTCAAAAAGATAGGGATATAAAATAACTTGGTGTCTGTGAATTTAATGAAGTGTCTC	176809
OY	305	TGTAGTGTAGCTTACATATAGAGACAGCAAGAAATTAATGAAGAGAAAGAGACATTTGAG	364
Db	176810	TGTAACTTATGTATATATAGATATGGGAATTAAGTGAAGAGAAAGAGAACATTTGAAG	176869
OY	365	CACCCAGAGAGAGAGGAGACCCTCCCTCTTAAAGTGAATGAGTGGCCTTCATTTCCAAGA	424
Db	176870	CACCCAGAGAGAGAGAGGGTCTCTCTCCCTTAAGTGAATGATGTGCCCTTTATCCGAGGA	176929
OY	425	GAAAGATTCAGGTGTCCCGCGGGAAGATGAGGGAGACCACATCCACAGGAATG-----	476
Db	176930	GAAAGATTCAGGTGTCCCTGTGGAAGAATGAGAGATCAATTTGTGAAGATGCCAGGAAG	176989
OY	477	-----GCAGAAATCAATCCTGTGTGCAAT	499
Db	176990	TCAATGCAAAAGTAATGATATCCACAGAGAAATGATATCCACAGAAATCATCCATGTGGCT	177049
OY	500	AAATGAGAGAGAGGGGCTCAAAAGATGAGCAAAAGAGATGAGCAAAAGAAATGGTGAATG	559
Db	177050	AAATGAGAGAGA-GGGGTCAAGGGGTGAGCAAAAGAGATGAAGCAAGTAAATCCTGTATGC	177108
OY	560	GGATACCTGAGAGATGGCCTGCTGTGTGTGAGCAAAATGTGGGCAAAATGGGCACTCCATG	619
Db	177109	AGATATCTGAGAGATGGCCCCAGCTGTATGATGAATAGGGTTTATGATATATGATGTAT	177168
OY	620	AAACAGACAGCTTGTCTGTGTTTTGCAGATCCTTAAATTAAGSCACATGGCATGCCATGAG	679
Db	177169	GGCAAGGCTGATGTGCCACAGGTGGCCAGATCCTTAATTAAGGGGGCAATTTGCAT-----	177222
OY	680	GCTAAGGGAGTGGAGGGGAAAGGATATATGATTAATGAGAGAAATGCAGAGAGAGCCAGA	739
Db	177223	GGAAGAGGCCACAGAGAGAGTGGAAAAAGAAAGTAATTTCAAAGATACAGAGAGAGCCAGA	177282
OY	740	AGGACAGGAGTAGAGAGGAGAGTTTGCACAGAGCTTTGTCTCTCCCAACAGCTCTCT	799
Db	177283	AGCAAAAGGATAGAGAGATGGGTTCAACAAAGCTTAACCTCCCACTA-----	177332
OY	800	CTTCCCTTCTATATGCATACATACACAGTGTGAGTGTGCAATGTGTGACATATGCAT	859
Db	177333	-----TGTGTATGAGATGCACAGCAATGTAGACTGTGGCTGTGCAATGTGCAATGTGGCT	177386
OY	860	GTGATGAACAGAGCCAGCTTGTGGTGTGCAAGTCTTTTCAGGCCCTATCTAATCTTTTGA	919
Db	177387	GTGAAGGCCAGAGGCCAATCTTGGGTGTCAATCTTCAGAGCACTATCTAATCTTGTCTGA	177446
OY	920	GACA-----ATCTCACTTGAAGT-----	940
Db	177447	GACACAGCTTCTACGTGGCTGATTCATTAAGCAGGTATGTGCTGTGCCAGCAGCAAGC	177506
OY	941	-----	940
Db	177507	CCCAGAGCTTGGCATGTATATGCTCCCTGTGTGGGATTTACAAAGCACACTATTCACAC	177566
OY	941	-----	940
Db	177567	CCAGCTTTTATGTGGGTGCCGGGATGGATTTGGTGACAAGTCTTTAAGCTTTGTGTGAC	177626
OY	941	-----TGAGTACCTCTCTAATTTCTACAGAGTTTCTCTACAGTGGGA	985
Db	177627	AGACACTTATGATGATAGTACTTACTTCCATCTCTACTACTAGAGTTCCTTCAAGTGGGGA	177686

QY	986	GGATGGGTGGGGAAGCAAAATTTAAAGCTGGTTGATTTCTTGAATTTCAAGTGGGGCTGG	1045
Db	177687	GGAAAGGGGTGGGAATAGCAAACTTTTGAAGATGGTTGGTTCTTGAATTTCAAGAGGGTTGG	177746
QY	1046	GAATATGACAGCTAATATATTCAGTTTCCTCGTTCCGGCTGGCTTCCCTGGGGTATCAAG	1105
Db	177747	GAATATGAACTATATATATTCAGTTTTCCTGGTTCCAGCTAGCTTTCCTGGGGGTATCAAC	177806
QY	1106	CAGAGTATAGTAGCCCTGTGTGGCAGTACACCAAGCAGACAGAAATAGGCAATGGCTC	1165
Db	177807	CAGTGAATAGTAGCCCTGTATGACAGCCACCAAGAGAACAGAGAGTGGGGTATGGCTC	177866
QY	1166	TGGTGGGCTGGTATGACATATAGAAAGATTCCTTGAAGCAAGATGTTGGCATCTCCAC	1225
Db	177867	TGGAGTGGTGGTATGACATATAGAAAGGGTTCCTATGATGCAAGATGTTTGTATCTTAG	177926
QY	1226	ACTATGACAGCCCGGAAAGTTTGGCTCCCGAGGACACCA-GCACTGAGACTGAGATG	1284
Db	177927	ACTTATGACAGCCCGGAAACTTTTGTCTCTTGAGGACACCGCAGCACTGAGACTGAGATG	177986
QY	1285	CATCAATTCAGAGACCAAGAAAGCAGGGTGTAGCACTTAGAAGAGACATAGCCAAA	1344
Db	177987	CATCAATTCAGAGAACCAAGAAACATATAGTGTCTAATAGCACTTAGAAGATAGACACTG	178046
QY	1345	GTCTCTTGTCT	1404
Db	178047	GTCTCTTGTCT	178106
QY	1405	TAAATTCAGAAAGGATACAAAGAAAGTCTCATGAGGACCTGTGTAAGAAAG-CATPAGG	1463
Db	178107	TAAATTCAGAGGAGATACAAAGAAATGAGATTTATCATGAGGACCTGTGTAAGAAAGCAG	178166
QY	1464	CATCAGTAATTAAGTTTACAAAGATTAACATAGTGGTGAATGTAAGAACAGACATGGCCA	1523
Db	178167	CATCAGTAATTAAGTTTACAAAGATTAACATAGTGGTGAATGTAAGAACAGACATGGCCA	178219
QY	1524	TGTTTTTTTGTATGAAACACACGACACAGGACGACGACACTCAGCTGTGGCGACGGCGC	1583
Db	178220	-----ACACATGTCATACACACATGTCACACATATATATATATATATATATATAT	178261
QY	1584	ACACACACACAGGCGACACACACACACACACACATGACATGACACACACACACACACATG	1643
Db	178262	ATGACACACACATGACACACACACACACACATGACACACATGATGACACATTAACATG	178321
QY	1644	AAAAGTGAATTAAGATATTTCTCACTTTGGCAAGTGGATGAAAGTTGATCAAAATG	1703
Db	178322	AAAAGTGAATTAAGATATTTCTCACTTTGACCAAGATGAAATGTTGGCCAAATG	178381
QY	1704	AAAGTATATCTCAGACTATTTTGTATGAGAGGAGTTATATAATTTATTTATTTAT	1763
Db	178382	AAAGTATATTTCAACATATTTTGTATGAGAGGAGTTATATAATTTATTTATTTAT	178441
QY	1764	ATATCTATTTACGTTTGTGGCAG-----CCATAGTTGGCTTGAATCTCACTA	1811
Db	178442	ATATCTATTTACGTTTGTGGCAGGGGTCTATGACACCTTAAGTTGGCTTCAAACTCACTA	178501
QY	1812	TGAAGTACGAATGACCTTGAGCTCTGATCCTTATATTTACCTCAAGTGGCCAGAT	1871
Db	178502	TGAAGTACGAATGACCTTGAGCTCTGATCCTTATATTTATTTATTTATTTATTTAT	178561
QY	1872	TATATAGTGCACACATATATCTCAGTTTATGCTGTGCTAAGCACTAAGCC	1921
Db	178562	TACAGGTGCTCTCCATATATCTCAGTTTATGCTGTGCTAAGCACTAAGCC	178611
RESULT 3			
AC11518		165503 bp	DNA
LOCUS			
DEFINITION			
AC11518			
VERSION			
KEYWORDS			
AC11518.2			
GI:21736068			
HTG; HTGS_PHASE1.			
HTG 23-Jul-2002			
SEQUENCING IN PROGRESS			

Wed Feb 19 14:09:08 2003

us-09-605-042a-1_copy_4000_6000.rge

SOURCE Rattus norvegicus.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 165503)
 Muzny, D.M., Adams, C., Adio-Oduola, B., Alt-osman, F.R., Allen, C.,
 Alstrooks, S.L., Amarantunge, H.C., Are, J.R., Ayele, M., Banks, T.,
 Barbata, D., Benton, J., Blmage, K., Blankenburg, K., Bonin, D.,
 Barbata, J., Bowie, S., Burkett, C., Burrell, K.L., Byrd, N.C.,
 Buhay, C., Burch, P., Burkett, C., Cavazos, S.R., Chacko, J., Chavez, D.,
 Caron, T.F., Carter, M., Chavaz, S., Chowdhry, I., Christopoulos, C.,
 Chen, G., Chen, R., Chen, Z., Coyle, M.D., Dathorne, S.R., David, R.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Delaney, K.R., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Douthett, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Eathwaite, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Galisti, A., Geo, J., Garcia, A., Garner, T., Hale, S., Hamilton, K.,
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hernandez, J.,
 Harris, C., Harris, K., Hart, M., Havlak, P., Hayes, A., Hollins, L.E.,
 Hernandez, O., Hodgson, A., Hogue, M., Hollway, C., Jackson, L.E.,
 Homi, F., Howard, S., Huber, J., Hulik, S., Hume, J., Joudah, S.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Kallison, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratoch, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louie, H.,
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maleshwar, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mahoney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabac, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenwo, S., Ogulu, M., Okunou, G.,
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peety, T., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Py, L.L., Quiles, M., Ren, Y.,
 Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Saverly, G.,
 Scherer, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I.,
 Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Tang, H.,
 Sutton, A., Svatek, A., Taber, P., Tameris, A., Tameris, K., Thomas, S.,
 Tansley, J., Taylor, C., Taylor, T., Tellrod, B., Vinson, N., Wang, Q.,
 Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, N., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watling, K.,
 Williams, G., Williamson, A., Wleczek, R., Woden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.

TITLE Unpublished
 JOURNAL Direct Submission
 REFERENCE 2 (bases 1 to 165503)
 AUTHORS Worley, K.C.
 JOURNAL Direct Submission
 TITLE Submitted (19-FEB-2002) Human Genome Sequencing Center, Department
 JOURNAL of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 165503)
 REFERENCE Direct Submission
 AUTHORS Worley, K.C.
 JOURNAL Submitted (23-JUL-2002) Human Genome Sequencing Center, Department
 TITLE of Molecular and Human Genetics, Baylor College of Medicine, One
 JOURNAL Baylor Plaza, Houston, TX 77030, USA
 COMMENT On Jul 12, 2002 this sequence version replaced gi:18701283.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GMD
 Center clone name: CH230-142F15
 ----- Summary Statistics
 Sequencing vector: Plasmid
 Chemistry: Dye-terminator Big Dye 100% of reads
 Assembly program: Phrap; version 0.990329

Consensus quality: 85091 bases at least Q40
 Consensus quality: 91433 bases at least Q30
 Consensus quality: 97328 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 74 contigs. The true order of the pieces
 * is not known and their order in this sequence recorded as
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated and the accession number will
 * be preserved.
 *
 * 1
 * 1037 1036: contig of 1036 bp in length
 * 1137 1136: gap of unknown length
 * 2147 1136: contig of 1010 bp in length
 * 2247 2246: gap of unknown length
 * 3339 3338: contig of 1092 bp in length
 * 3439 3438: gap of unknown length
 * 5182 5181: contig of 1743 bp in length
 * 5282 5281: gap of unknown length
 * 6562 6561: gap of 1280 bp in length
 * 6662 6661: gap of unknown length
 * 7949 7949: contig of 1288 bp in length
 * 7950 7949: gap of unknown length
 * 8050 8049: gap of 1057 bp in length
 * 9107 9106: contig of 1057 bp in length
 * 9207 9206: gap of unknown length
 * 10569 10568: contig of 1362 bp in length
 * 10669 10668: gap of unknown length
 * 11970 11969: contig of 1301 bp in length
 * 12070 12069: gap of 1191 bp in length
 * 13261 13260: contig of unknown length
 * 13361 13360: gap of 1084 bp in length
 * 14445 14444: contig of unknown length
 * 14445 14444: gap of unknown length
 * 14445 14444: contig of 1027 bp in length
 * 15571 15570: gap of unknown length
 * 15672 15671: gap of 1174 bp in length
 * 15752 15751: contig of 1174 bp in length
 * 16845 16844: gap of unknown length
 * 16945 16944: gap of 1151 bp in length
 * 16946 16945: contig of unknown length
 * 18097 18096: gap of unknown length
 * 18197 18196: gap of 1156 bp in length
 * 19353 19352: contig of unknown length
 * 19453 19452: gap of 1427 bp in length
 * 20879 20878: contig of unknown length
 * 20880 20879: gap of unknown length
 * 22248 22247: gap of 1269 bp in length
 * 22349 22348: gap of unknown length
 * 22349 22348: gap of 1261 bp in length
 * 23609 23608: contig of 1261 bp in length
 * 23709 23708: gap of unknown length
 * 23710 23709: gap of 1254 bp in length
 * 24964 24963: contig of 1254 bp in length
 * 25064 25063: gap of unknown length
 * 26601 26600: gap of 1537 bp in length
 * 26701 26700: gap of unknown length
 * 27732 27731: gap of 1031 bp in length
 * 27732 27731: contig of unknown length
 * 27732 27731: gap of 1227 bp in length
 * 29058 29057: contig of 1227 bp in length
 * 29058 29057: gap of unknown length
 * 29159 29158: contig of 1076 bp in length
 * 30235 30234: gap of unknown length
 * 30235 30234: gap of 1344 bp in length
 * 31678 31677: gap of unknown length
 * 31679 31678: gap of 1600 bp in length
 * 33379 33378: contig of unknown length
 * 33379 33378: gap of unknown length
 * 33479 33478: gap of 1236 bp in length
 * 34715 34714: contig of unknown length
 * 34715 34714: gap of unknown length
 * 34815 34814: gap of 1365 bp in length
 * 36180 36179: contig of 1438 bp in length
 * 36280 36279: gap of unknown length
 * 37718 37717: contig of 1438 bp in length
 * 37818 37817: gap of unknown length
 * 39087 39086: contig of 1269 bp in length
 * 39186 39185: gap of unknown length
 * 40741 40740: contig of 1555 bp in length

```

*      40742      40841: gap of unknown length
*      40842      42312: contig of 1471 bp in length
*      42313      42412: gap of unknown length
*      42413      43475: contig of 1063 bp in length
*      43476      43575: gap of unknown length
*      43576      45143: contig of 1568 bp in length
*      45144      45243: gap of unknown length
*      45244      46483: contig of 1240 bp in length
*      46484      46583: gap of unknown length
*      46584      48590: contig of 2007 bp in length
*      48591      48690: gap of unknown length
*      48691      50392: contig of 1702 bp in length
*      50393      50492: gap of unknown length
*      50493      52222: contig of 1730 bp in length
*      52223      53322: gap of unknown length
*      53323      53891: contig of 1569 bp in length
*      53892      55698: contig of 1707 bp in length
*      55699      55798: gap of unknown length
*      55799      57953: contig of 2155 bp in length
*      57954      58053: gap of unknown length
*      58054      59390: contig of 1357 bp in length
*      59391      59490: gap of unknown length
*      59491      61859: contig of 2369 bp in length
*      61860      61959: gap of unknown length
*      61960      63467: contig of 1508 bp in length
*      63468      63567: gap of unknown length
*      63568      65055: contig of 1488 bp in length
*      65056      65156: gap of unknown length
*      65157      66516: contig of 1361 bp in length
*      66517      68266: gap of unknown length
*      68267      68726: contig of 2010 bp in length
*      68727      70038: gap of unknown length
*      70039      70138: contig of 1312 bp in length
*      70139      71379: gap of unknown length
*      71380      71479: contig of 1241 bp in length
*      71480      73457: gap of unknown length
*      73458      73557: contig of 1978 bp in length
*      73558      75747: gap of unknown length
*      75748      75847: contig of 2190 bp in length
*      75848      77681: gap of unknown length
*      77682      77781: contig of 1834 bp in length
*      77782      79051: gap of unknown length
*      79052      79151: contig of 1270 bp in length
*      79152      79559: gap of unknown length
*      81580      81679: contig of 2428 bp in length
*      81680      84013: gap of unknown length
*      84013      84013: contig of 2334 bp in length

```

```

Query Match      30.9% Score 617.6; DB 2; Length 165503;
Best Local Similarity 80.4%; Pred. No. 2.3e-157; Indels 40; Gaps 5;
Matches 800; Conservative 0; Mismatches 155;

```

```

OY 941 TGAGTGAAGCTCTAGTATTTCTACAGAGCTTCTCAGTGGGAGAGAAATGGGTGGGAGA 1000
DB 33541 TGAGTGAAGCTCTAGTATTTCTACAGAGCTTCTCAGTGGGAGAGAAATGGGTGGGAGA 1000
OY 1001 ACCAAATTTAGACTGGTGTGATTTCTTCATTAATTCAGTGGGTGGGAAATAGCACTATA 1060
DB 33601 ACCAAATTTAGACTGGTGTGATTTCTTCATTAATTCAGTGGGTGGGAAATAGCACTATA 1060
OY 1061 TATTCAAGTTTCTCTGCTCTCTGCTGCTCTCTGCTGCTCTCTGCTGCTCTCTGCTGCT 33720
DB 33661 TATTCAAGTTTCTCTGCTCTCTGCTGCTCTCTGCTGCTCTCTGCTGCTCTCTGCTGCT 33720
OY 1121 CTGTGTGGAGTCAACAGACAGACAGAAAGATGGGCTGCTGCTGCTGCTGCTGCTGCTG 33780
DB 33721 CTGTGTGGAGTCAACAGACAGACAGAAAGATGGGCTGCTGCTGCTGCTGCTGCTGCTG 33780
OY 1181 ACATAGGAAGATCTTGTGAGAGATGTTGCTCATCTCAGAGACTTGAAGACCCAG 1240
DB 33781 ACATAGGAAGATCTTGTGAGAGATGTTGCTCATCTCAGAGACTTGAAGACCCAG 1240

```

```

OY 1241 GAAAGTTTCTCTCCAGGACAGCCAGCA-GCAGTGAAGTGAATGCATCAATCCAGAGA 1299
DB 33841 GAAAGTTTCTCTCTCCAGGACAGCCAGCCAGTGAAGTGAATGCATCAATCCAGAGA 1299
OY 1300 CCAGAAAGACAGGCTCTGACATAGAGAGACATAGCCCAAAAGTCTCTCTCTCTG 33960
DB 33901 CCAGAAAGACAGGCTCTGACATAGAGAGACATAGCCCAAAAGTCTCTCTCTCTG 33960
OY 1360 CCTAAGCTTTGGCATTTCTGCAAACTTGAAAATTAGCATCTTTAATTCAGAGAGA 1419
DB 33961 CCTAAGCTTTGGCATTTCTGCAAACTTGAAAATTAGCATCTTTAATTCAGAGAGA 1419
OY 1420 TACAGAGAGAGACTTACATGAGGACCTTGTAAGAGAGACAGGAGTAAAGTAAAG 34080
DB 34021 TACAGAGAGAGACTTACATGAGGACCTTGTAAGAGAGAGACAGGAGTAAAGTAAAG 34080
OY 1479 TTCAAGATATACATCATAGTGTGAGTGAACAAAGGACATGCTTTTGTGTAT 1538
DB 34081 TTCAAGATATACATCATAGTGTGAGTGAACAAAGGACATGCTTTTGTGTAT 1538
OY 1539 GAAACACAGCAGCAGGACAGGACCTGCTGCGCAGCGCCGACACACACACACGCG 34175
DB 34081 TTCAAGAGAGAGATCATGATAGTAAACAAAGTAAAC 34118
OY 1599 CACACACACACACACGATGACACATGACACACACACACAACTGCAAAAGTAAATA 1658
DB 34119 CACACACACACACACGATGACACATGACACACACACACAACTGCAAAAGTAAATA 1658
OY 1659 GATATTTCTGCTTTGGCAAGTGTGATGAAAGTGTATGTAATTAATTAATTAATTA 1718
DB 34176 GATATTTCTGCTTTGGCAAGTGTGATGAAAGTGTATGTAATTAATTAATTAATTA 1718
OY 1719 ACTATTTGTACTAGAGGAGGATTAATTAATTAATTAATTAATTAATTAATTAATTA 1778
DB 34236 ACTATTTGTACTAGAGGAGGATTAATTAATTAATTAATTAATTAATTAATTAATTA 1778
OY 1779 TTTGGTCAG-----CCTAAGTGGCTTGAACCTACATGAGCTAGCTAGCA 1826
DB 34296 TTTGGTCAG-----CCTAAGTGGCTTGAACCTACATGAGCTAGCTAGCA 1826
OY 1827 CTTGAGCTCTGATCTTATATCTATCTACACTGCAAGTGGCCAGATTATTAATGCTACCA 1886
DB 34356 CTTGAGCTCTGATCTTATATCTATCTACACTGCAAGTGGCCAGATTATTAATGCTACCA 1886
OY 1887 CTATACCTAGTTTATGCTGTGCTGCTAGGACTAGCC 1921
DB 34416 CTATACCTAGTTTATGCTGTGCTGCTAGGACTAGCC 1921
OY 1887 CTATACCTAGTTTATGCTGTGCTGCTAGGACTAGCC 34509
DB 34475 CTATACCTAGTTTATGCTGTGCTGCTAGGACTAGCC 34509

```

```

RESULT 4
ALB31768/c 156954 bp DNA linear HTG 17-AUG-2002
LOCUS Dantio rerio clone RP71-101, *** SEQUENCING IN PROGRESS ***, 6
DEFINITION
unoriented pieces.
ACCESSION
ALB31768.5 GI:22416172
VERSION
HTG: HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS
zebrafish.
SOURCE
Dantio rerio
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Dantio.
1 (bases 1 to 156954)
REFERENCE
Sinha, H.
Direct Submission
Wellcome Trust Sanger Institute, Hinxton,
Submitted (16-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridge, CB10 1SA, UK. E-mail enquiries: zface@sanger.ac.uk
Cambridge, CB10 1SA, UK. E-mail enquiries: zface@sanger.ac.uk
On Aug 21, 2002 this sequence version replaced gi:22204775.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk

```

[illegible]

```

* 29674 29773: gap of 100 bp
* 29774 65168: contig of 35396 bp in length
* 65170 65269: gap of 100 bp
* 65270 78612: contig of 13343 bp in length
* 78613 78712: gap of 100 bp
* 78713 94994: contig of 16282 bp in length
* 94995 95094: gap of 100 bp
* 95095 98073: contig of 2879 bp in length
* 98074 98173: gap of 100 bp
* 98174 154679: contig of 56506 bp in length
* 154680 154779: gap of 100 bp
* 154780 168153: contig of 13374 bp in length.
FEATURES
source
  Location/Qualifiers
    1..168153
      /organism="Danio rerio"
      /db_xref="taxon:7955"
      /clone_lib="CH211-8A9"
      /clone_1lb="CHORI-211"
      1..6555
        /note="assembly-fragment:01614
        fragment_chain:1"
        misc_feature
          /note="assembly-fragment:01045
          fragment_chain:1"
          29774..65169
            /note="assembly-fragment:02181
            fragment_chain:1"
            misc_feature
              /note="assembly-fragment:01821
              fragment_chain:1"
              78713..94994
                /note="assembly-fragment:00604
                fragment_chain:2"
                misc_feature
                  /note="assembly-fragment:02093
                  fragment_chain:2"
                  98174..154679
                    /note="assembly-fragment:00940
                    fragment_chain:3"
                    misc_feature
                      /note="assembly-fragment:00155
                      fragment_chain:3"
BASE COUNT  55901 a 29002 c 28390 g 54160 t 700 others
ORIGIN
Query Match 4.9%; Score 97.4; DB 2; Length 168153;
Best Local Similarity 50.3%; Pred. No. 1.2e-15;
Matches 239; Conservative 0; Mismatches 236; Indels 0; Gaps 0;
QY 1527 TTTTGTGTTAAGAACACACAGCAGGACGACTCAGGTGCGCAGCGGCACACA 1586
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 14431 TATATATATATACACACACACACACACACACACACACACACACACACACACA 14430
QY 1587 CACACACAGCGCACACACACACACACACACACACACACACACACACACACACACA 1646
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 14491 CACACACACACACACACACACACACACACACACACACACACACATATTATACATTTTGGCA 14550
QY 1647 AGTGATATATATATATATATATATATATATATATATATATATATATATATATATAT 1706
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 14551 ATTATATATATATATATATATATATATATATATATATATATATATATATATATAT 14610
QY 1707 GTTATATACAGACATATTTGTATAGAGGAGGTATTAATTTGTATATTTATATAT 1766
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 14611 AACAAATTAATTTTAAACAATTTAAACATGCACATTTACAGCTACAGTAAGAAAGATAT 14670
QY 1767 CTATTTACTGTTTGGCAGGCTAGGTGCTGTGAACACACATATAGCTAGCAATGA 1826
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 14671 CATGCTACTTTTCTGCAGGGGTTAACCTAGTCCAGGGCTTGTCAGACGTCATCATAG 14730
QY 1827 CATTGAGCTTCTGATCCTTATATATATATATATATATATATATATATATATATATAT 1886
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 14731 TCTTTGCTTCTGTCAGCCACATATTTCAATTGAAGCGCAAAATTCGCGTCCACCAAT 14790

```

```

QY 1887 CTATCTCAGTTTATCTGTCTGCTPAGGACTPAGCCCAATTATACAAACACACACATAT 1946
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 14791 TCTGCGACTGTGGGCCGAAACACTGAGCTGAGACACACACACACACACACACACAC 14850
QY 1947 ATACACACATACACACACACACACACACATATATATATATATATATATATATATACACA 2001
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 14851 ACACACACACACACACACACACACACATATATACATATACATPACACGACATATACA 14905
RESULT 6
AC068493/c
LOCUS AC068493
DEFINITION Mus musculus clone RP23-76K1 strain C57BL6/J, WORKING DRAFT
SEQUENCE, 26 unordered pieces.
AC068493
VERSION AC068493.10 GI:15148081
KEYWORDS HTGS, HTGS_PHASE1, HTGS_DRAFT.
SOURCE Mus musculus.
ORGANISM Mus musculus.
REFERENCE 1
AUTHORS Montgomery,K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,
Toshkhes,I.P., Shim,C., Decker,J., Thomas,E., Ferreira,A.,
Gordon,M., Goltz,J.S., and Kuchelapatti,R.
JOURNAL High Throughput Mouse Sequencing
REFERENCE 2 (bases 1 to 269081)
AUTHORS Montgomery,K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,
Toshkhes,I.P., Shim,C., Decker,J., Thomas,E., Ferreira,A.,
Gordon,M., Goltz,J.S., and Kuchelapatti,R.
JOURNAL Direct Submission
REFERENCE 3 (03-MAY-2000) Department of Molecular Genetics, Albert
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
Bronx, NY 10461, USA
COMMENT On Aug 11, 2001 this sequence version replaced gi:1493654.
-----Genome Center
Center: Harvard Partners Genome Center
Center Code: HPGC
Web site: http://www.hpcg.org/Sequence/mouse.html
Contact: hpgc@mdel.mgh.harvard.edu
-----Summary Statistics
Center project name: ABN
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 100%
*Consensus quality: 256302 at least Q20
*Consensus quality: 254124 at least Q30
*Consensus quality: 250276 at least Q40
*Estimated insert size: agarose-FP - N/A
**Estimated insert size: 268581 - sum-of-coverage
Quality coverage: agarose-FP - N/A
Quality coverage: 6.2 x 1n Q20 bases; sum-of-coverage estimation
-----
NOTE: This is a "working draft" sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
51730: contig of 51730 bp in length
51731 51750: gap of unknown length
51751 82423: contig of 30673 bp in length
82424 82443: gap of unknown length
82444 118158: contig of 35715 bp in length
118159 118178: gap of unknown length
118179 139441: contig of 21263 bp in length
139442 139462: gap of unknown length
139463 160638: contig of 21177 bp in length
160639 160658: gap of unknown length
160659 177611: contig of 16953 bp in length
177612 177631: gap of unknown length

```

*	177632	180991:	contig of 13360 bp in length
*	1909392	191011:	gap of unknown length
*	191012	200128:	contig of 9117 bp in length
*	200129	200148:	gap of unknown length
*	200149	211882:	contig of 11734 bp in length
*	211883	211902:	gap of unknown length
*	211903	220773:	contig of 8811 bp in length
*	220794	220944:	gap of unknown length
*	220794	229664:	contig of 8850 bp in length
*	229644	229663:	gap of unknown length
*	229664	238339:	contig of 8616 bp in length
*	238340	238359:	gap of unknown length
*	238360	244848:	contig of 6489 bp in length
*	244849	244866:	gap of unknown length
*	244869	249723:	contig of 4857 bp in length
*	249726	249745:	gap of unknown length
*	249746	252523:	contig of 2178 bp in length
*	252524	252543:	gap of unknown length
*	252544	257520:	contig of 4977 bp in length
*	257521	257540:	gap of unknown length
*	257541	260396:	contig of 2856 bp in length
*	260397	260416:	gap of unknown length
*	260417	262709:	contig of 2293 bp in length
*	262710	262728:	gap of unknown length
*	262730	263934:	contig of 1205 bp in length
*	263935	263934:	gap of unknown length
*	263935	264196:	contig of 242 bp in length
*	264197	264216:	gap of unknown length
*	264217	264629:	contig of 413 bp in length
*	264630	264640:	gap of unknown length
*	264650	265608:	contig of 959 bp in length
*	265609	265628:	gap of unknown length
*	265629	266933:	contig of 1307 bp in length
*	266936	266955:	gap of unknown length
*	266956	268333:	contig of 1380 bp in length
*	268336	268335:	gap of unknown length
*	268350	268545:	contig of 194 bp in length
*	268550	268561:	gap of unknown length
*	268570	269081:	contig of 512 bp in length.

FEATURES
SOURCE

misc_feature	1. .51730
misc_feature	/note="assembly_name:Contig138"
misc_feature	51751. .82423
misc_feature	/note="assembly_name:Contig137"
misc_feature	82444. .118158
misc_feature	/note="assembly_name:Contig136"
misc_feature	118179. .139441
misc_feature	/note="assembly_name:Contig135"
misc_feature	139462. .160638
misc_feature	/note="assembly_name:Contig134"
misc_feature	160659. .177611
misc_feature	/note="assembly_name:Contig133"
misc_feature	177633. .190991
misc_feature	/note="assembly_name:Contig132"
misc_feature	191012. .200128
misc_feature	/note="assembly_name:Contig131"
misc_feature	200149. .211882
misc_feature	/note="assembly_name:Contig130"
misc_feature	211903. .220773
misc_feature	/note="assembly_name:Contig129"
misc_feature	220794. .229643
misc_feature	/note="assembly_name:Contig128"
misc_feature	229664. .238339
misc_feature	/note="assembly_name:Contig127"
misc_feature	238360. .244848
misc_feature	/note="assembly_name:Contig126"
misc_feature	244869. .249725

```
misc_feature /note="assembly_name:Contig125"
249746..252523 /note="assembly_name:Contig124
clone_end:SP6
vector_side:left"
252544..257520 /note="assembly_name:Contig123"
257541..260396 /note="assembly_name:Contig122"
260417..262709 /note="assembly_name:Contig121"
262730..263934 /note="assembly_name:Contig120"
263955..264196 /note="assembly_name:Contig119"
264217..264629 /note="assembly_name:Contig118"
264650..265608 /note="assembly_name:Contig117"
265629..266935 /note="assembly_name:Contig116"
266956..268335 /note="assembly_name:Contig115"
268356..268549 /note="assembly_name:Contig114"
268570..269081 /note="assembly_name:Contig113"
/note="assembly_name:Contig112"
269102..269394 /note="assembly_name:Contig111"
269415..269707 /note="assembly_name:Contig110"
269728..270020 /note="assembly_name:Contig109"
270041..270333 /note="assembly_name:Contig108"
270354..270646 /note="assembly_name:Contig107"
270667..270959 /note="assembly_name:Contig106"
270980..271272 /note="assembly_name:Contig105"
271293..271585 /note="assembly_name:Contig104"
271606..271898 /note="assembly_name:Contig103"
271919..272211 /note="assembly_name:Contig102"
272232..272524 /note="assembly_name:Contig101"
272545..272837 /note="assembly_name:Contig100"
272858..273150 /note="assembly_name:Contig99"
273171..273463 /note="assembly_name:Contig98"
273484..273776 /note="assembly_name:Contig97"
273797..274089 /note="assembly_name:Contig96"
274110..274402 /note="assembly_name:Contig95"
274423..274715 /note="assembly_name:Contig94"
274736..275028 /note="assembly_name:Contig93"
275049..275341 /note="assembly_name:Contig92"
275362..275654 /note="assembly_name:Contig91"
275675..275967 /note="assembly_name:Contig90"
275988..276280 /note="assembly_name:Contig89"
276301..276593 /note="assembly_name:Contig88"
276614..276906 /note="assembly_name:Contig87"
276927..277219 /note="assembly_name:Contig86"
277240..277532 /note="assembly_name:Contig85"
277553..277845 /note="assembly_name:Contig84"
277866..278158 /note="assembly_name:Contig83"
278179..278471 /note="assembly_name:Contig82"
278492..278784 /note="assembly_name:Contig81"
278805..279097 /note="assembly_name:Contig80"
279118..279410 /note="assembly_name:Contig79"
279431..279723 /note="assembly_name:Contig78"
279744..280036 /note="assembly_name:Contig77"
280057..280349 /note="assembly_name:Contig76"
280370..280662 /note="assembly_name:Contig75"
280683..280975 /note="assembly_name:Contig74"
280996..281288 /note="assembly_name:Contig73"
281309..281601 /note="assembly_name:Contig72"
281622..281914 /note="assembly_name:Contig71"
281935..282227 /note="assembly_name:Contig70"
282248..282540 /note="assembly_name:Contig69"
282561..282853 /note="assembly_name:Contig68"
282874..283166 /note="assembly_name:Contig67"
283187..283479 /note="assembly_name:Contig66"
283500..283792 /note="assembly_name:Contig65"
283813..284105 /note="assembly_name:Contig64"
284126..284418 /note="assembly_name:Contig63"
284439..284731 /note="assembly_name:Contig62"
284752..285044 /note="assembly_name:Contig61"
285065..285357 /note="assembly_name:Contig60"
285378..285670 /note="assembly_name:Contig59"
285691..285983 /note="assembly_name:Contig58"
285996..286288 /note="assembly_name:Contig57"
286301..286593 /note="assembly_name:Contig56"
286614..286906 /note="assembly_name:Contig55"
286927..287219 /note="assembly_name:Contig54"
287240..287532 /note="assembly_name:Contig53"
287553..287845 /note="assembly_name:Contig52"
287866..288158 /note="assembly_name:Contig51"
288179..288471 /note="assembly_name:Contig50"
288492..288784 /note="assembly_name:Contig49"
288805..289097 /note="assembly_name:Contig48"
289118..289410 /note="assembly_name:Contig47"
289431..289723 /note="assembly_name:Contig46"
289744..290036 /note="assembly_name:Contig45"
290057..290349 /note="assembly_name:Contig44"
290370..290662 /note="assembly_name:Contig43"
290683..290975 /note="assembly_name:Contig42"
290996..291288 /note="assembly_name:Contig41"
291309..291601 /note="assembly_name:Contig40"
291622..291914 /note="assembly_name:Contig39"
291935..292227 /note="assembly_name:Contig38"
292248..292540 /note="assembly_name:Contig37"
292561..292853 /note="assembly_name:Contig36"
292874..293166 /note="assembly_name:Contig35"
293187..293479 /note="assembly_name:Contig34"
293500..293792 /note="assembly_name:Contig33"
293813..294105 /note="assembly_name:Contig32"
294126..294418 /note="assembly_name:Contig31"
294439..294731 /note="assembly_name:Contig30"
294752..295044 /note="assembly_name:Contig29"
295065..295357 /note="assembly_name:Contig28"
295378..295670 /note="assembly_name:Contig27"
295691..295983 /note="assembly_name:Contig26"
295996..296288 /note="assembly_name:Contig25"
296301..296593 /note="assembly_name:Contig24"
296614..296906 /note="assembly_name:Contig23"
296927..297219 /note="assembly_name:Contig22"
297240..297532 /note="assembly_name:Contig21"
297553..297845 /note="assembly_name:Contig20"
297866..298158 /note="assembly_name:Contig19"
298179..298471 /note="assembly_name:Contig18"
298492..298784 /note="assembly_name:Contig17"
298805..299097 /note="assembly_name:Contig16"
299118..299410 /note="assembly_name:Contig15"
299431..299723 /note="assembly_name:Contig14"
299744..300036 /note="assembly_name:Contig13"
300057..300349 /note="assembly_name:Contig12"
300370..300662 /note="assembly_name:Contig11"
300683..300975 /note="assembly_name:Contig10"
300996..301288 /note="assembly_name:Contig9"
301309..301601 /note="assembly_name:Contig8"
301622..301914 /note="assembly_name:Contig7"
301935..302227 /note="assembly_name:Contig6"
302248..302540 /note="assembly_name:Contig5"
302561..302853 /note="assembly_name:Contig4"
302874..303166 /note="assembly_name:Contig3"
303187..303479 /note="assembly_name:Contig2"
303500..303792 /note="assembly_name:Contig1"
303813..304105 /note="assembly_name:Contig0"
304126..304418 /note="assembly_name:Contig-1"
304439..304731 /note="assembly_name:Contig-2"
304752..305044 /note="assembly_name:Contig-3"
305065..305357 /note="assembly_name:Contig-4"
305378..305670 /note="assembly_name:Contig-5"
305691..305983 /note="assembly_name:Contig-6"
305996..306288 /note="assembly_name:Contig-7"
306301..306593 /note="assembly_name:Contig-8"
306614..306906 /note="assembly_name:Contig-9"
306927..307219 /note="assembly_name:Contig-10"
307240..307532 /note="assembly_name:Contig-11"
307553..307845 /note="assembly_name:Contig-12"
307866..308158 /note="assembly_name:Contig-13"
308179..308471 /note="assembly_name:Contig-14"
308492..308784 /note="assembly_name:Contig-15"
308805..309097 /note="assembly_name:Contig-16"
309118..309410 /note="assembly_name:Contig-17"
309431..309723 /note="assembly_name:Contig-18"
309744..310036 /note="assembly_name:Contig-19"
310057..310349 /note="assembly_name:Contig-20"
310370..310662 /note="assembly_name:Contig-21"
310683..310975 /note="assembly_name:Contig-22"
310996..311288 /note="assembly_name:Contig-23"
311309..311601 /note="assembly_name:Contig-24"
311622..311914 /note="assembly_name:Contig-25"
311935..312227 /note="assembly_name:Contig-26"
312248..312540 /note="assembly_name:Contig-27"
312561..312853 /note="assembly_name:Contig-28"
312874..313166 /note="assembly_name:Contig-29"
313187..313479 /note="assembly_name:Contig-30"
313500..313792 /note="assembly_name:Contig-31"
313813..314105 /note="assembly_name:Contig-32"
314126..314418 /note="assembly_name:Contig-33"
314439..314731 /note="assembly_name:Contig-34"
314752..315044 /note="assembly_name:Contig-35"
315065..315357 /note="assembly_name:Contig-36"
315378..315670 /note
```

[illegible]

VERSION AC123466.2 GI:21908309
 KEYWORDS HTG: PHASE1.
 SOURCE Rattus norvegicus.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE

AUTHORS

1 (bases 1 to 179047)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T.,
 Babadad,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
 Bouck,J., Bowles,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
 Bunay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
 Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisedge,H.,
 Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
 Massey,E., McWhiney,E., McLeod,M.P., Meador,M., Mel,G., Metker,M.,
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okunou,G.,
 Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
 Rives,M., Rojas,A., Rojudoan,I., Rolfe,M., Ruiz,S., Saverly,G.,
 Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,
 Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Swatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
 Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
 Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstein,G. and Gibbs,R.

TITLE

JOURNAL

2 (bases 1 to 179047)

Worley,K.C.

Direct Submission

Submitted (29-MAY-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 179047)

Worley,K.C.

Direct Submission

Submitted (23-JUL-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Jul 19, 2002 this sequence version replaced gi:21240407.

COMMENT

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GXPJ

Center clone name: CH230-18J2

Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 131140 bases at least Q40
 Consensus quality: 137974 bases at least Q30
 Consensus quality: 143605 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 NOTE: This is a 'working draft' sequence. It currently
 consists of 67 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1	1260:	contig of 1260 bp in length
1261	1360:	gap of unknown length
1361	2596:	contig of 1236 bp in length
2597	2696:	gap of unknown length
2697	3723:	contig of 1033 bp in length
3730	3823:	gap of unknown length
3830	5306:	contig of 1477 bp in length
5307	5406:	gap of unknown length
5407	7000:	contig of 1594 bp in length
7001	7100:	gap of unknown length
7101	8265:	contig of 1165 bp in length
8266	8365:	gap of unknown length
8366	9628:	contig of 1263 bp in length
9629	9728:	gap of unknown length
9729	10850:	contig of 1122 bp in length
10851	10950:	gap of unknown length
10951	12094:	contig of 1144 bp in length
12095	12194:	gap of unknown length
12195	13347:	contig of 1153 bp in length
13348	13447:	gap of unknown length
13448	14876:	contig of 1429 bp in length
14877	14976:	gap of unknown length
14977	16027:	contig of 1051 bp in length
16028	16127:	gap of unknown length
16128	17519:	contig of 1392 bp in length
17520	17619:	gap of unknown length
17620	19103:	contig of 1490 bp in length
19110	19209:	gap of unknown length
19210	20619:	contig of 1410 bp in length
20620	20719:	gap of unknown length
20720	22755:	contig of 2036 bp in length
22756	22855:	gap of unknown length
22856	23985:	contig of 1130 bp in length
23986	24085:	gap of unknown length
24086	25490:	contig of 1405 bp in length
25491	25590:	gap of unknown length
25591	26609:	contig of 1019 bp in length
26610	26709:	gap of unknown length
26710	27744:	contig of 1035 bp in length
27745	27844:	gap of unknown length
27845	29057:	contig of 1213 bp in length
29058	29157:	gap of unknown length
29158	30413:	contig of 1256 bp in length
30414	30513:	gap of unknown length
30514	32203:	contig of 1596 bp in length
32210	32309:	gap of unknown length
32310	33716:	contig of 1407 bp in length
33717	33816:	gap of unknown length
33817	35558:	contig of 1742 bp in length
35559	35658:	gap of unknown length
35659	37534:	contig of 1876 bp in length
37535	37634:	gap of unknown length
37635	39211:	contig of 1577 bp in length
39212	39311:	gap of unknown length
39312	41424:	contig of 2113 bp in length
41425	41524:	gap of unknown length
41525	42771:	contig of 1247 bp in length

Wed Feb 19 14:09:08 2003

us-09-605-042a-1_copy_4000_6000.rge

42772	42871: gap of unknown length
42872	42871: contig of 1205 bp in length
44077	44176: gap of unknown length
44077	44176: contig of 1840 bp in length
46017	46116: gap of unknown length
46017	46116: contig of 3261 bp in length
49378	49377: gap of unknown length
49378	49377: contig of 1728 bp in length
49478	51205: gap of unknown length
51206	51305: gap of 2409 bp in length
51306	53714: gap of unknown length
53715	53814: gap of 1742 bp in length
53815	55556: gap of unknown length
55557	55556: contig of 2406 bp in length
55657	58062: gap of unknown length
58063	58162: gap of 2273 bp in length
60336	60435: gap of unknown length
60336	60435: contig of 3310 bp in length
63846	63945: gap of unknown length
63846	63945: contig of 2325 bp in length
66271	66270: gap of unknown length
66271	66270: contig of 1526 bp in length
67897	67896: gap of unknown length
67897	67896: contig of 2224 bp in length
70221	70220: gap of unknown length
70221	70220: contig of 2259 bp in length
72579	72579: gap of unknown length
72580	72579: contig of 2627 bp in length
73507	73506: gap of unknown length
73507	73506: contig of 2652 bp in length
78159	78158: gap of unknown length
78159	78158: contig of 3175 bp in length
81334	81333: gap of unknown length
81334	81333: contig of 1548 bp in length
82981	82981: gap of unknown length
82982	82981: contig of 3831 bp in length
83082	86912: gap of unknown length
83082	86912: contig of 2687 bp in length
86913	86999: gap of unknown length
86913	86999: contig of 2345 bp in length
89700	92144: gap of unknown length
89700	92144: contig of 1789 bp in length
92145	94033: gap of unknown length
92145	94033: contig of 3317 bp in length
94034	97450: gap of unknown length
94034	97450: contig of 2785 bp in length
97451	100335: gap of unknown length
97451	100335: contig of 2452 bp in length
100336	102887: gap of unknown length
100336	102887: contig of 2943 bp in length
102888	105930: gap of unknown length
102888	105930: contig of 179047;

Query Match 4.3%; Score 86.6; DB 2; Length 179047;
Best Local Similarity 61.1%; Pred. No. 1.1e-12; Indels 1; Gaps 1;
Matches 157; Conservative 0; Mismatches 99;

Db 90535 ACATCATTAATGAAAAATA 90551

RESULT 8	166494/c	166494	7218 bp	DNA	linear	PAT 28-DEC-1997
LOCUS	Sequence 14 from patent US 5670367.					
DEFINITION	Sequence 14 from patent US 5670367.					
ACCESSION	166494					
VERSION	166494.1	GI:2724471				
KEYWORDS						
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	1 (bases 1 to 7218)					
AUTHORS	Donner, F., Scheiflinger, F. and Falkner, F. Gunter.					
TITLE	Recombinant fowlpox virus					
JOURNAL	US 5670367-A 14 23-SEP-1997;					
FEATURES	Location/Qualifiers					
source	1..7218					
BASE COUNT	1944 a 1491 c 1486 g 1929 t					368 others
ORIGIN						

Query Match 4.3%; Score 86; DB 6; Length 7218;
Best Local Similarity 5.5%; Pred. No. 1.5e-12; Indels 0; Gaps 0;
Matches 23; Conservative 249; Mismatches 144;

Db 374	AGAGAGGACCTCTCCCTMAAGTGAATGAGTGGCTCTCATTCAGAGAGAGAGATT	1396
Db 1455	AGAGATGAGAAATTTGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR	493
Db 434	CAGCTGCCCGGAGATGAGGACCAATCCACAGGAATGGCGAGATCATCTGT	1336
Db 1395	RR	553
Db 494	GTGCATAATGAGAGAGGCGGCTCAAGATGAGCAAGAAGATGACGAAGAATGCT	1276
Db 1335	RR	613
Db 554	GGATGTGATCTCTGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1216
Db 1275	RR	673
Db 614	TCATGACACAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1156
Db 1215	RR	733
Db 674	ATGAGAGCTAGGAGAGTGGAGGGAAGTATAGATGATGACAGAGTACAGAGAG	1096
Db 1155	RR	789
Db 734	CCAGAGAGACAGAGAGTGGAGGAGAGGCTTGTGCTGCTGCTGCTGCTGCTGCT	1040
Db 1095	RR	

RESULT 9	AC122205	227271 bp	DNA	linear	HTG 23-MAY-2002
LOCUS	Mus musculus chromosome UNK clone RP23-76c13, WORKING DRAFT				
DEFINITION	SEQUENCE, 49 unordered pieces.				
ACCESSION	AC122205.1	GI:21105035			
VERSION	AC122205.1	GI:21105035			
KEYWORDS	HTGS, PHASE1, HTGS, DRAFT.				
SOURCE	Mus musculus.				
ORGANISM	Mus musculus.				
REFERENCE	1 (bases 1 to 227271)				
AUTHORS	McPherson, J.D. and Waterston, R.H.				
TITLE	The sequence of Mus musculus clone				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 227271)				

AUTHORS McPherson, J.D. and Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (23-MAY-2002) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA

COMMENT

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc/index.shtml
 Contact: submissions@wustl.wustl.edu
 Project Information
 Center project name: M_BA0076C13

 Summary Statistics
 Sequencing vector: M13; 0%
 Sequencing vector: plasmid; 100%
 Chemistry: Dye-terminator ET; 0% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 213736 bases at least Q40
 Consensus quality: 217681 bases at least Q30
 Consensus quality: 220647 bases at least Q20
 Insert size: 219000; agarose-fp
 Insert size: 235763; sum-of-contigs
 Quality coverage: 3.82 in Q20 bases; agarose-fp
 Quality coverage: 3.44 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 49 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 1 1281: contig of 1281 bp in length
 * 1282 1381: gap of unknown length
 * 1382 2691: contig of 1310 bp in length
 * 2692 2791: gap of unknown length
 * 2792 4003: contig of 1212 bp in length
 * 4004 4104: gap of unknown length
 * 4104 5381: contig of 1278 bp in length
 * 5382 5481: gap of unknown length
 * 5482 6625: contig of 1144 bp in length
 * 6626 6725: gap of unknown length
 * 6726 8003: contig of 1278 bp in length
 * 8004 8103: gap of unknown length
 * 8104 10089: contig of 1986 bp in length
 * 10090 10189: gap of unknown length
 * 10190 11462: contig of 1273 bp in length
 * 11463 11562: gap of unknown length
 * 11563 12633: contig of 1071 bp in length
 * 12634 12733: gap of unknown length
 * 12734 13913: contig of 1180 bp in length
 * 13914 14013: gap of unknown length
 * 14014 16055: contig of 2042 bp in length
 * 16056 16155: gap of unknown length
 * 16156 17394: contig of 1239 bp in length
 * 17395 17494: gap of unknown length
 * 17495 18889: contig of 1395 bp in length
 * 18890 18990: gap of unknown length
 * 18990 20703: contig of 1714 bp in length
 * 20704 20803: gap of unknown length
 * 20804 23097: contig of 2294 bp in length
 * 23098 23197: gap of unknown length
 * 23198 25368: contig of 2171 bp in length
 * 25369 25468: gap of unknown length
 * 25469 28468: contig of 3000 bp in length
 * 28469 28568: gap of unknown length
 * 28569 30438: contig of 1770 bp in length
 * 30439 32905: gap of unknown length
 * 32906 33005: contig of 2467 bp in length
 * 33005: gap of unknown length

33006 35141: contig of 2136 bp in length
 * 35142 35241: gap of unknown length
 * 35242 38554: contig of 3313 bp in length
 * 38555 38654: gap of unknown length
 * 38655 41056: contig of 2402 bp in length
 * 41057 41157: gap of unknown length
 * 41157 43820: contig of 2664 bp in length
 * 43821 43920: gap of unknown length
 * 43921 47161: contig of 3241 bp in length
 * 47162 47261: gap of unknown length
 * 47262 49573: contig of 2312 bp in length
 * 49574 49673: gap of unknown length
 * 49674 53370: contig of 3697 bp in length
 * 53371 53470: gap of unknown length
 * 53471 55725: contig of 2255 bp in length
 * 55726 55825: gap of unknown length
 * 55826 59849: contig of 4024 bp in length
 * 59850 59950: gap of unknown length
 * 59950 64330: contig of 4381 bp in length
 * 64331 64431: gap of unknown length
 * 64431 70187: contig of 5757 bp in length
 * 70188 70287: gap of unknown length
 * 70288 75140: contig of 4853 bp in length
 * 75141 75240: gap of unknown length
 * 75241 79513: contig of 4273 bp in length
 * 79514 79613: gap of unknown length
 * 79614 83717: contig of 4104 bp in length
 * 83718 83817: gap of unknown length
 * 83818 88509: contig of 4652 bp in length
 * 88510 88609: gap of unknown length
 * 88610 93754: contig of 5145 bp in length
 * 93755 93854: gap of unknown length
 * 93855 98912: contig of 5058 bp in length
 * 98913 99012: gap of unknown length
 * 99013 103620: contig of 4608 bp in length
 * 103621 103720: gap of unknown length
 * 103721 109365: contig of 5645 bp in length
 * 109366 109465: gap of unknown length
 * 109466 115864: contig of 6399 bp in length
 * 115865 115964: gap of unknown length
 * 115965 124808: contig of 8844 bp in length
 * 124809 124908: gap of unknown length
 * 124909 131207: contig of 6299 bp in length
 * 131208 131307: gap of unknown length
 * 131308 139527: contig of 8220 bp in length
 * 139528 139627: gap of unknown length
 * 139628 147766: contig of 8139 bp in length
 * 147767 147866: gap of unknown length
 * 147867 160060: contig of 12194 bp in length
 * 160061 160160: gap of unknown length
 * 160161 171302: contig of 11142 bp in length
 * 171303 171402: gap of unknown length
 * 171403 184792: contig of 13390 bp in length
 * 184793 184892: gap of unknown length
 * 184893 199072: contig of 14180 bp in length
 * 199073 199172: gap of unknown length
 * 199173 213427: contig of 14155 bp in length
 * 213428 227271: contig of 13844 bp in length.
 * 213428

FEATURES

Source

1..227271
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="GNK"
 /clone="Rp23-76C13"
 misc_feature
 1..1281
 /note="assembly_name:Contig52"
 misc_feature
 1382..2691
 /note="assembly_name:Contig76"
 misc_feature
 2792..4003
 /note="assembly_name:Contig1"
 misc_feature
 4104..5381
 /note="assembly_name:Contig82"

```

misc_feature      5482..6625
                  /note="assembly_name:Contig85"
misc_feature      6726..8003
                  /note="assembly_name:Contig86"
misc_feature      8104..10089
                  /note="assembly_name:Contig88"
misc_feature      10190..11462
                  /note="assembly_name:Contig89"
misc_feature      11563..12633
                  /note="assembly_name:Contig90"
misc_feature      12734..13913
                  /note="assembly_name:Contig91"
misc_feature      14014..16055
                  /note="assembly_name:Contig92"
misc_feature      16156..17394
                  /note="assembly_name:Contig93"
misc_feature      17495..18889
                  /note="assembly_name:Contig94"
misc_feature      18990..20703
                  /note="assembly_name:Contig95"
misc_feature      20804..23097
                  /note="assembly_name:Contig97"
misc_feature      23198..25368
                  /note="assembly_name:Contig98"
misc_feature      25469..28468
                  /note="assembly_name:Contig99"
misc_feature      28569..30338
                  /note="assembly_name:Contig100"
misc_feature      30439..32905
                  /note="assembly_name:Contig101"
misc_feature      33006..35141
                  /note="assembly_name:Contig102"
misc_feature      35242..38554
                  /note="assembly_name:Contig103"
misc_feature      38655..41056
                  /note="assembly_name:Contig104"
misc_feature      41157..43820
                  /note="assembly_name:Contig105"
misc_feature      43921..47161

```

Query Match Best Local Similarity 4.0%; Score 79.8; DB 2; Length 227271;

Matches 131; Conservative 0; Mismatches 67; Indels 1; Gaps 1;

```

QY 1804 ACTGACTATGAGCTAGCAATGACCTTGAGCTTGATCTATATCTTACA-CTCTCAAG 1862
      ||||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 137223 ATTTCTATGTTAGTCCAAAGGACCTGAGCTCTTGACCTTTGCGCTTACCTGCCGAG 137164

QY 1863 TGCCAGATTATAGTGTGACACACTATCTCACTTATGCTGTGCTAAGACTAAGCCC 1922
      ||| ||||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 137163 TGCTTGATTTATAGGACCCACCACTACCTGTTGTGCTAGCGTGGGCTGGAACCT 137104

QY 1923 AATTACAAACACACATATATACACATACACACACACACACACATATATAT 1982
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 137103 TCATACATACATACATACATACATACATACATACATACATACATACATACAT 137044

QY 1983 GTATATATATATACATACA 2001
      ||| ||| ||| ||| |||
Db 137043 ACATACATACATACATACA 137025

```

```

RESULT 10
AC100738/c      123116 bp      DNA      1linear      HTG 20-AUG-2002
LOCUS          Mus musculus clone RP24-362K21, WORKING DRAFT SEQUENCE, 2 unordered
DEFINITION     pieces:
AC100738
VERSION        AC100738.2 GI:22325091
KEYWORDS       HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FUZZTOP.
SOURCE         house mouse.
ORGANISM       Mus musculus.
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

REFERENCE

1 (bases 1 to 123116)
 Birren,B., Nusbaum,C. and Lander,E.
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 Unpublished
 2 (bases 1 to 123116)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,A., Allen,N.,
 Anderson,S., Barina,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
 Brown,A., Camarata,J., Campolano,A., Chang,J., Chazaro,B.,
 Choepel,Y., Colangelo,M., Collins,S., Collins,A., Cook,A.,
 Cooke,P., Dearielano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
 Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
 Hago,B., Hearford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
 Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K.,
 Lamazares,R., Landers,T., Lenockky,J., Levine,R., Liu,G.,
 Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
 McCarthy,M., McEwan,P., McKernan,K., McSheeters,R., Meldrum,J.,
 Menus,L., Mihova,T., Mlenka,V., Murphy,T., Naylor,J., Nguyen,C.,
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
 Oliver,J., Peterson,K., Phunhang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Riback,M., Riley,R., Rise,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnpack,R.,
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zalnoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE

JOURNAL
 REFERENCE
 AUTHORS
 Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 123116)

TITLE

JOURNAL
 REFERENCE
 AUTHORS
 Submitted (20-AUG-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 20, 2002 this sequence version replaced gi:17048105.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WITR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center clone name: 362-K-21
 Summary Statistics
 Sequencing vector: plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 123002 bases at least Q40
 Consensus quality: 123006 bases at least Q30
 Consensus quality: 123011 bases at least Q20
 Insert size: 115000; agarose-gel
 Insert size: 123016; sum-of-contigs


```

* 5671 6386: contig of 716 bp in length
* 6387 6486: gap of 100 bp
* 6487 7207: contig of 721 bp in length
* 7208 7307: gap of 100 bp
* 7308 8047: contig of 740 bp in length
* 8048 8147: gap of 100 bp
* 8148 8874: contig of 727 bp in length
* 8875 8974: gap of 100 bp
* 8975 9697: contig of 723 bp in length
* 9698 9797: gap of 100 bp
* 9798 10537: contig of 740 bp in length
* 10538 10637: gap of 100 bp
* 10638 11362: contig of 725 bp in length
* 11363 11462: gap of 100 bp
* 11463 12162: contig of 700 bp in length
* 12163 12262: gap of 100 bp
* 12263 12965: contig of 703 bp in length
* 12966 13065: gap of 100 bp
* 13066 13781: contig of 716 bp in length
* 13782 13881: gap of 100 bp
* 13882 14579: contig of 698 bp in length
* 14580 14679: gap of 100 bp
* 14680 15381: contig of 702 bp in length
* 15382 15481: gap of 100 bp
* 15482 16224: contig of 743 bp in length
* 16225 16324: gap of 100 bp
* 16325 17042: contig of 718 bp in length
* 17043 17142: gap of 100 bp
* 17143 17897: contig of 755 bp in length
* 17898 17997: gap of 100 bp
* 17998 18738: contig of 741 bp in length
* 18739 18838: gap of 100 bp
* 18839 19574: contig of 736 bp in length
* 19575 19674: gap of 100 bp
* 19675 20381: contig of 707 bp in length
* 20382 20481: gap of 100 bp
* 20482 21186: contig of 705 bp in length
* 21187 21286: gap of 100 bp
* 21287 22004: contig of 718 bp in length
* 22005 22104: gap of 100 bp
* 22105 22823: contig of 719 bp in length
* 22824 22923: gap of 100 bp
* 22924 23641: contig of 718 bp in length
* 23642 23741: gap of 100 bp
* 23742 24458: contig of 717 bp in length
* 24459 24558: gap of 100 bp
* 24559 25293: contig of 735 bp in length
* 25294 25393: gap of 100 bp
* 25394 26115: contig of 722 bp in length
* 26116 26215: gap of 100 bp
* 26216 26935: contig of 720 bp in length
* 26936 27035: gap of 100 bp
* 27036 27755: contig of 720 bp in length
* 27756 27855: gap of 100 bp
* 27856 28570: contig of 715 bp in length
* 28571 28670: gap of 100 bp
* 28671 29376: contig of 706 bp in length
* 29377 29476: gap of 100 bp
* 29477 30211: contig of 735 bp in length
* 30212 30311: gap of 100 bp
* 30312 31007: contig of 696 bp in length
* 31008 31107: gap of 100 bp
* 31108 31830: contig of 723 bp in length
* 31831 31930: gap of 100 bp
* 31931 32653: contig of 723 bp in length
* 32654 32753: gap of 100 bp
* 32754 33485: contig of 732 bp in length
* 33486 33585: gap of 100 bp
* 33586 34290: contig of 705 bp in length
* 34291 34390: gap of 100 bp
* 34391 35139: contig of 749 bp in length
* 35140 35239: gap of 100 bp
* 35240 35942: contig of 703 bp in length

```

```

* 35943 36042: gap of 100 bp
* 36043 36788: contig of 746 bp in length
* 36789 36888: gap of 100 bp
* 36889 37620: contig of 732 bp in length
* 37621 37720: gap of 100 bp
* 37721 38447: contig of 727 bp in length
* 38448 38547: gap of 100 bp
* 38548 39250: contig of 703 bp in length
* 39251 39350: gap of 100 bp
* 39351 40050: contig of 700 bp in length
* 40051 40150: gap of 100 bp
* 40151 40844: contig of 694 bp in length
* 40845 40944: gap of 100 bp
* 40945 41661: contig of 717 bp in length
* 41662 41761: gap of 100 bp
* 41762 42487: contig of 726 bp in length
* 42488 42587: gap of 100 bp
* 42588 43316: contig of 729 bp in length
* 43317 43416: gap of 100 bp
* 43417 44138: contig of 722 bp in length
* 44139 44238: gap of 100 bp
* 44239 44957: contig of 719 bp in length
* 44958 45057: gap of 100 bp
* 45058 45798: contig of 741 bp in length
* 45799 45898: gap of 100 bp
* 45899 46620: contig of 722 bp in length
* 46621 46720: gap of 100 bp
* 46721 47439: contig of 719 bp in length
* 47440 47539: gap of 100 bp
* 47540 48267: contig of 728 bp in length
* 48268 48367: gap of 100 bp
* 48368 49078: contig of 711 bp in length
* 49079 49178: gap of 100 bp
* 49179 49887: contig of 709 bp in length
* 49888 49987: gap of 100 bp
* 49988 50703: contig of 716 bp in length
* 50704 50803: gap of 100 bp
* 50804 51534: contig of 731 bp in length
* 51535 51634: gap of 100 bp
* 51635 52372: contig of 738 bp in length
* 52373 52472: gap of 100 bp
* 52473 53226: contig of 754 bp in length
* 53227 53326: gap of 100 bp
* 53327 54057: contig of 731 bp in length
* 54058 54157: gap of 100 bp
* 54158 54860: contig of 703 bp in length
* 54861 54960: gap of 100 bp
* 54961 55687: contig of 727 bp in length
* 55688 55787: gap of 100 bp
* 55788 56488: contig of 701 bp in length
* 56489 56588: gap of 100 bp

Query Match      3.8%; Score 76.8; DB 2; Length 70420;
Best Local Similarity 59.0%; Pred. No. 4.9e-10;
Matches 151; Conservative 0; Mismatches 102; Indels 3; Gaps 1;

QY 1749 TATTGTTATGTTATATCTATTTTACTGTTGTGGCACCCTTAAGTGGCTTGAACTCA 1808
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24653 TTTTGTGTTGTTGTTGCTTAAGCAAAAGTCATGATCCAGGATGGATCCAACTAG 24712

QY 1809 CTTAGAGCTAGCAATGACCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1865
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24713 ACATGTAGCCAGATGACCTGGAATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 24772

QY 1866 CCAGATTATAGTGTGACCACTATATCTAGTTTATGCTGCTGAAGCAAT 1925
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24773 TAGATTAACAGCGCTGTGACCAACATGCTGATTTATAGTACGGGATGAACCAAG 24832

QY 1926 TATACAAACACACACATATATACACATATACACACACACACACACACATATATATGTA 1985
      || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 24833 GTTTCATTCACACACACACACACACACACACACACACACACACACACACACACACACAC 24892

QY 1986 TATATATATACATACA 2001

```

Db 24893 CACACACACACACCA 24908

RESULT 12
AC128306/c

LOCUS

DEFINITION AC128306 145883 bp DNA linear HTG 19-JUL-2002
Rattus norvegicus clone CH230-247A20.*** SEQUENCING IN PROGRESS
*** 49 unordered pieces.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

1 (bases 1 to 145883)
Muzny, D.M., Adams, C., Adio-Ogunola, B., Alt-Osman, F.R., Allen, C.,
Alstbrooks, S.L., Amaralunga, H.C., Are, J.R., Ayale, M., Banks, T.,
Barbata, J., Benton, J., Bimaga, K., Blankenburg, K., Bonin, D.,
Bouck, J., Bowle, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P.,
Buhay, C., Burck, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhury, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dethorne, S.R., David, R.,
DeVella, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorrell, J.H., Guevara, M., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hognes, M., Holloway, C., Hollins, B.,
Hosni, F., Howard, S., Huber, J., Huliy, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
Karlssoon, E., Kelly, S., Khan, O., King, L., Korvan, J., Kovar, C.,
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louisedge, H.,
Lozdo, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M.,
Miner, G., Miner, Z., Mitchell, T., Monabath, R., Morgan, M., Morris, S.,
Mosier, M., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokwenkwo, S., Ogulu, M., Okunolu, G.,
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
Rivers, M., Rojas, A., Rojudoan, I., Rolfe, M., Ruiz, S., Savery, G.,
Scherer, S., Scott, G., Shen, H., Shooshitari, N., Sisson, I.,
Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Tabor, P., Tameris, A., Tameris, K., Tang, H.,
Tansley, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Usmani, K., Vasquez, L., Vera, V., Villalobos, D., Vinson, R., Wang, H.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wleczek, R., Woodson, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G., and Gibbs, R.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

COMMENT

Unpublished
2 (bases 1 to 145883)
Worley, K.C.
Direct Submission
Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GZGE

Center clone name: CH230-247A20
----- Summary Statistics

Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 109118 bases at least Q40
Consensus quality: 113028 bases at least Q30
Consensus quality: 116051 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 49 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1	1016:	contig of 1016 bp in length
*	1017	1116: gap of unknown length
*	1117	2133: contig of 1017 bp in length
*	2134	2233: gap of unknown length
*	2234	3436: contig of 1203 bp in length
*	3437	3536: gap of unknown length
*	3537	4775: contig of 1239 bp in length
*	4776	4875: gap of unknown length
*	4876	6085: contig of 1210 bp in length
*	6086	6185: gap of unknown length
*	6186	7934: contig of 1749 bp in length
*	7935	8034: gap of unknown length
*	8035	9328: contig of 1294 bp in length
*	9329	9428: gap of unknown length
*	9429	10821: contig of 1303 bp in length
*	10832	10921: gap of unknown length
*	10922	12241: contig of 1320 bp in length
*	12242	12341: gap of unknown length
*	12342	13419: contig of 1078 bp in length
*	13420	13519: gap of unknown length
*	13520	15432: contig of 1913 bp in length
*	15433	15532: gap of unknown length
*	15533	17561: contig of 2029 bp in length
*	17562	17661: gap of unknown length
*	17662	18446: contig of 1185 bp in length
*	18447	18946: gap of unknown length
*	18947	20852: contig of 1906 bp in length
*	20853	20952: gap of unknown length
*	20953	22110: contig of 1158 bp in length
*	22111	22210: gap of unknown length
*	22211	24244: contig of 2034 bp in length
*	24245	24344: gap of unknown length
*	24345	26564: contig of 2220 bp in length
*	26565	26664: gap of unknown length
*	26665	28667: contig of 2203 bp in length
*	28668	28967: gap of unknown length
*	28968	31186: contig of 2219 bp in length
*	31187	31286: gap of unknown length
*	31287	34321: contig of 3035 bp in length
*	34322	34421: gap of unknown length
*	34422	36921: contig of 2500 bp in length
*	36922	37021: gap of unknown length
*	37022	39379: contig of 2358 bp in length
*	39380	39479: gap of unknown length
*	39480	41820: contig of 2341 bp in length
*	41821	41920: gap of unknown length
*	41921	44885: contig of 2965 bp in length
*	44886	44985: gap of unknown length
*	44986	47233: contig of 2248 bp in length
*	47234	47333: gap of unknown length
*	47334	48337: contig of 2504 bp in length
*	48338	49937: gap of unknown length
*	49938	52598: contig of 2661 bp in length
*	52599	52698: gap of unknown length

```

* 52699 54894: contig of 2196 bp in length
* 54895 54994: gap of unknown length
* 54995 57159: contig of 2165 bp in length
* 57160 57259: gap of unknown length
* 57260 61271: contig of 4012 bp in length
* 61272 61371: gap of unknown length
* 61372 65576: contig of 4205 bp in length
* 65577 65676: gap of unknown length
* 65677 68706: contig of 3030 bp in length
* 68707 68806: gap of unknown length
* 68807 72844: contig of 4038 bp in length
* 72845 72944: gap of unknown length
* 72945 75709: contig of 2765 bp in length
* 75710 75809: gap of unknown length
* 75810 79442: contig of 3633 bp in length
* 79443 79542: gap of unknown length
* 79543 82979: contig of 3437 bp in length
* 82980 83079: gap of unknown length
* 83080 85611: contig of 2532 bp in length
* 85612 85711: gap of unknown length
* 85712 90083: contig of 4372 bp in length
* 90084 90183: gap of unknown length
* 90184 93473: contig of 3290 bp in length
* 93474 93573: gap of unknown length
* 93574 96658: contig of 3085 bp in length
* 96659 96758: gap of unknown length
* 96759 100861: contig of 4103 bp in length
* 100862 100961: gap of unknown length
* 100962 105785: contig of 4824 bp in length
* 105786 105885: gap of unknown length
* 105886 110787: contig of 4902 bp in length
* 110788 110887: gap of unknown length
* 110888 115196: contig of 4309 bp in length
* 115197 115296: gap of unknown length
* 115297 121380: contig of 6084 bp in length
* 121381 121480: gap of unknown length
* 121481 127794: contig of 6314 bp in length
* 127795 127894: gap of unknown length
* 127895 131144: contig of 3250 bp in length
* 131145 131244: gap of unknown length
* 131245 138374: contig of 7130 bp in length
* 138375 138474: gap of unknown length
* 138475 145883: contig of 7409 bp in length.
* 145883 Location/Qualifiers
  source          1. 145883
                  /organism="Rattus norvegicus"
                  /db_xref="taxon:10116"
                  /clone="CH230-247A20"
BASE COUNT      39927 a 31216 c 31537 g 38313 t 4890 others
ORIGIN
Query Match      3.8%; Score 75.8; DB 2; Length 145883;
Best Local Similarity 52.0%; Pred. No. 9.3e-10;
Matches 193; Conservative 0; Mismatches 177; Indels 1; Gaps 1;
OY 1540 AACACAGCAGGAGGACGACGTCAGCGGCGGCGGACACACACACGCGC 1599
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 72474 ATACACACAAACATACATACATACATACATACATACATACATACATAC 72415
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1600 ACACACACACACACGATCAGACATGAC-CACACACAAACTGCAAAAGTAA 1658
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 72414 ACATACATATATATACACACATATATACATACATACATACATACAT 72355
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1659 GATATTTCTGACTTTGGCAAAGTGCATGCAAAAGTTGCAAAATGAAAGT 1718
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 72354 ACATGACCTTCCCTGTCTGCTAGCTGCTGCTATATCATCTCATCTG 72295
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1719 ACTATTTGTACTAGAGGAGGTTAAATATGTTATGTTATTTCTATTTACT 1778
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 72294 TATTTATTTATTTATTTGTTGTTGTTGTTGTTGTTGTTATTTATTT 72235
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1779 TTGTGCGACGCTTAAGTTGCTCTGGAAGTCACTATGAGAGTGAAGCTT 1838
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

DB 72234 CTCAGGATTTTCATGTTGGCTTCACACTACATACATCTAGGGGATCTGA 72175
OY 1839 GATCCTATATCTACACTGTCACATGAGCCACATATAGTGACACCATATAC 1898
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 72174 GGCTTCCTGCTGTATGCAAGAGTGCTGGGATTTACAGGATGTGCCAC 72115
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1899 TATGCTGTGCT 1909
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 72114 CATATCATGCT 72104

RESULT 13
AL845338/c
LOCUS
DEFINITION
  Mus musculus chromosome 4 clone RP24-480C2, *** SEQUENCING IN
  PROGRESS ***, 8 unordered pieces.
ACCESSION
  AL845338
VERSION
  AL845338.2 GI:22265601
KEYWORDS
  HTG; HTGS; PHASE1.
SOURCE
  house mouse.
  Mus musculus.
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 (bases 1 to 172121)
REFERENCE
  Submitted (11-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
  Cambridgeshire, CB10 1SA UK. E-mail enquiries:
  humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
  On Aug 15, 2002 this sequence version replaced gi:22205033.
  ----- Genome Center
  Center: UK Medical Research Council
  Center code: UK-MRC
  Web site: http://mrcseq.har.mrc.ac.uk
  Contact: mouse@har.mrc.ac.uk
  ----- Project Information
  Center project name: BM480C2
  ----- Summary Statistics
  Assembly program: XGAP4; Version 4.5
  Chemistry: Dye-terminator; 100% of reads
  Consensus quality: 167988 bases at least Q40
  Consensus quality: 169562 bases at least Q30
  Consensus quality: 170400 bases at least Q20
  Insert size: 171421; sum-of-contigs
  Insert size: 161713; 4.7% error; agarose-1p
  Quality coverage: 5.08x in Q20 bases; sum-of-contigs Quality
  Coverage: 5.42x in Q20 bases; agarose-1p
  -----
  * NOTE: This is a 'working draft' sequence. It currently
  * consists of 8 contigs. The true order of the pieces
  * is not known and their order in this sequence record is
  * arbitrary. Gaps between the contigs are represented as
  * runs of N, but the exact sizes of the gaps are unknown.
  * This record will be updated with the finished sequence
  * as soon as it is available and the accession number will
  * be preserved.
  *
  1 47162: contig of 47162 bp in length
  * 47163 47262: gap of 100 bp
  * 47263 64639: contig of 17377 bp in length
  * 64640 64739: gap of 100 bp
  * 64740 131541: contig of 66802 bp in length
  * 131542 131641: gap of 100 bp
  * 131642 138993: contig of 7352 bp in length
  * 138994 139093: gap of 100 bp
  * 139094 143007: contig of 3914 bp in length
  * 143008 143107: gap of 100 bp
  * 143108 148923: contig of 5816 bp in length
  * 148924 149023: gap of 100 bp
  * 149024 164521: contig of 15498 bp in length
  * 164522 164621: gap of 100 bp
  * 164622 172121: contig of 7500 bp in length.
  *
  Location/Qualifiers
    1. 172121

```

```

BASE COUNT      52910  a  32844  c  32754  g  52899  t
ORIGIN
misc_feature     /organism="Mus musculus"
                  /db_xref="taxon:10090"
                  /chromosome="4"
                  /clone="RP24-480C2"
                  /clone_1lb="RPci-24"
                  1. .477162
misc_feature     /note="assembly_fragment:01710
                  fragment_chain:1
                  clone_end:SP6
                  vector_side:left"
                  472633..664639
misc_feature     /note="assembly_fragment:00505
                  fragment_chain:1"
                  64740..131541
misc_feature     /note="assembly_fragment:01765
                  fragment_chain:1"
                  131642..138993
misc_feature     /note="assembly_fragment:00564
                  fragment_chain:1"
                  133094..143007
misc_feature     /note="assembly_fragment:00337
                  fragment_chain:1"
                  143108..148933
misc_feature     /note="assembly_fragment:00450
                  fragment_chain:1"
                  14024..164521
misc_feature     /note="assembly_fragment:00258
                  fragment_chain:1"
                  164622..172121
misc_feature     /note="assembly_fragment:01048
                  fragment_chain:1"

```

Query Match	3.8%;	Score 75.6;	DB 2;	Length 172121;
Best Similarity	53.0%;	Pred. No. 1.1e-09;		
Matches 229; Conservative	0;	Mismatches 199;	Indels 4;	Gaps 3;

OY	1573	GCGACGGGCGCACAGACACAGCGGCACACACACACAGCATGCACACAC	1632
Dd	69372	CACT	69311
OY	1633	ACACAACCTGCAAAAGTGAATAAAAAGATATTTCACATTTGGCAAAGTGANAAGT	1692
Dd	69312	TTCACATGTCTTCTCCTCGTCTGGTATTCTTGTCTGATTAGTAATAACAAG-AAGT	6925
OY	1693	TGATCAAAATGAAAGTATATACACAGACATTTTGTACTAGAGGGATTATTAATTAT	1752
Dd	69253	CTATTTTTAAGCTGTATTGAGAAAAGTTCTGTATTAGTAGATGAAATGCTATGTC	69194
OY	1753	GTTATTGTATAT--TCTATTTTACGTGTTGGCAGCCTAAGTTGGTCTGACACACT	1810
Dd	69193	TGTGAAGTCATTTGATCTACCTTGTGTTTATTTCTTCATTTGATTAACAATATTA	69133
OY	1811	ATGAAGCTAGCAATGACCCTTGAGCTTCTGATCCCTATATCTACACTCTCAAGTCCCCAGA	1870
Dd	69133	AACAAGGCAAGAAAGGTACTGAAGGCTTAACTCTGCTGCTGAATCAGAACCTATCTGAC	69074
OY	1871	TTATAGTGTGCACACTATFACTCAAGTTATCTGTGCTGAAGGACCTAACCC-CAATTATA	1929
Dd	69073	TTTTTATTGTTTCATTAATTTAGAACTCCTAACATTGGGTACACACACACACACACACA	69014
OY	1930	CAAAACACACACATATATACACACATACACACACACACACACAGATATATATATATATAT	1989
Dd	69013	CACACACACACACACACACACACACACATATATATATATATATATATATATATATATATA	6895
OY	1990	TATATACATACA 2001	
Dd	68953	TATATATATATA 68942	

LOCUS	AC124397	228421 bp	DNA	linear	HTG 21-Jul-2002
DEFINITION	Mus musculus chromosome DUK clone RP24-348J16, WORKING DRAFT				
ACCESSION	AC124397				
VERSION	AC124397.2 GI:21913423				
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FILLTOP.				
SOURCE	Mus musculus.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	1 (bases 1 to 228421)				
TITLE	McPherson,J.D. and Waterston,R.H.				
JOURNAL	The sequence of Mus musculus clone				
REFERENCE	Unpublished				
AUTHORS	2 (bases 1 to 228421)				
TITLE	McPherson,J.D. and Waterston,R.H.				
JOURNAL	Direct Submission				
REFERENCE	Submitted (14-JUN-2002) Genome Sequencing Center, 4444 Forest Park				
AUTHORS	Parkway, St. Louis, MO 63108, USA				
TITLE	3 (bases 1 to 228421)				
JOURNAL	McPherson,J.D. and Waterston,R.H.				
REFERENCE	Direct Submission				
AUTHORS	Submitted (21-JUL-2002) Genome Sequencing Center, 4444 Forest Park				
TITLE	Parkway, St. Louis, MO 63108, USA				
JOURNAL	On Jul 21, 2002 this sequence version replaced gi:21426516.				
COMMENT					

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Web site: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
Contact: submissions@wustl.edu
----- Project Information -----
Center project name: M_BB0348J16

```

----- Summary Statistics -----
Sequencing vector: M13, 0%
Sequencing vector: plasmid: 100%
Chemistry: Dye-primer: Err: 0% of reads
Chemistry: Dye-terminator: Big Dye: 100% of reads
Assembly program: Phrap: version 0.990319
Consensus quality: 22614 bases at least Q40
Consensus quality: 22667 bases at least Q30
Consensus quality: 226920 bases at least Q20
Insert size: 197000; average=fp
Insert size: 230199; sum-of-configs

```

Quality coverage: 17.61 in Q20 bases; agarose-fp
Quality coverage: 12.04 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*	1	2210:	contig of 2210 bp in length
*	2211	gap of unknown length	
*	2311	74701:	contig of 72391 bp in length
*	74702	74801:	gap of unknown length
*	74803	145379:	contig of 70938 bp in length
*	145740	145833:	gap of unknown length
*	145840	227901:	contig of 82062 bp in length
*	227902	228001:	gap of unknown length
*	228003	228095:	contig of 49 bp in length
*	228051	gap of unknown length	
*	228150	228215:	contig of 271 bp in length.

FEATURES

RESULT 14
AC124397/c

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP24-348J16"

```


RESULT	15
AL844571	
LOCUS	180964 bp DNA linear HTG 24-JUL-2007
DEFINITION	Mus musculus chromosome 2 clone RP23-224C23, *** SEQUENCING IN
PROGRESS	***, 7 unordered pieces.
AL844571	
ACCESSION	AL844571.1 GI:21955783
VERSION	HTG; HTGS_PHASE1.
KEYWORDS	Mus musculus.
SOURCE	Mus musculus.
ORGANISM	Mus musculus
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 180964)
AUTHORS	Plumb, B.
TITLE	Direct Submission
JOURNAL	Submitted (23-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,

```

COMMENT
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bm224c23
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 179438 bases at least Q40
Consensus quality: 179854 bases at least Q40
Consensus quality: 180125 bases at least Q20
Insert size: 180364; sum-of-contigs
Insert size: 172737; 3.4% error; agarose-fp
Quality coverage: 5.95x in Q20 bases; sum-of-contigs quality
coverage: 6.73x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.

```

```

misc_feature      /clone_lib="RPCI-23"
                  1..87139
/note="assembly_fragment:01784
clone_end:T7
vector_side:left"
87240..98867
/note="assembly_fragment:01085
fragment_chain:1"
98968..148932
/note="assembly_fragment:00972
fragment_chain:1"
149033..151927
/note="assembly_fragment:00380
fragment_chain:1"
152028..155409
/note="assembly_fragment:01200
fragment_chain:1"
155510..158708
/note="assembly_fragment:00473
fragment_chain:1"
158809..180964
/note="assembly_fragment:01345
fragment_chain:1
clone_end:SP6
vector_side:right"

```

[illegible]

[illegible]

Search completed: February 17, 2003, 15:04:36
Job time : 7849.09 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 17, 2003, 08:54:07 ; Search time 50.4863 Seconds
(without alignments)
9111.676 Million cell updates/sec

Title: US-09-605-042a-1_COPY_1_1500

Perfect score: 1500
Sequence: 1 gggggggggccctcgagggtt.....agatcccccattctctcg 1500

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued_Patents_NA.*
2: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/Backfilest1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	DB ID	Description
1	74	4.9	7218	1	US-08-232-463-14
2	40.8	2.7	42571	4	US-09-810-347-3
3	35.6	2.4	43950	4	US-09-735-934A-3
4	35.4	2.4	1166	1	US-08-599-252-95
5	35.4	2.4	1166	5	PCF-US96-06352-95
6	35.4	2.4	1166	5	PCF-US96-06583-95
7	35.4	2.4	5537	1	US-08-135-511-32
8	35.4	2.4	5537	1	US-08-483-852-9
9	35.4	2.4	5537	1	US-08-361-458-4
10	35.4	2.4	5537	1	US-08-477-953-9
11	35.4	2.4	5537	1	US-08-187-453-32
12	35.4	2.4	5537	1	US-08-562-985A-5
13	35.4	2.4	5537	2	US-08-477-952-9
14	35.4	2.4	8758	4	US-09-799-345-3
15	33.6	2.2	1339	4	US-09-442-143A-10
16	33.6	2.2	5403	4	US-09-442-143A-3
17	33.4	2.2	277	4	US-09-007-005-3
18	33.4	2.2	277	4	US-09-244-796-3
19	33.4	2.2	48974	4	US-08-920-422-17
20	33.2	2.2	35828	4	US-09-449-218D-17
21	33	2.2	2117	1	US-08-431-080-23
22	33	2.2	2117	2	US-08-938-534-23
23	33	2.2	2117	4	US-09-345-294-23
24	32.6	2.2	248	4	US-09-007-005-32
25	32.6	2.2	248	4	US-09-244-796-32
26	32.6	2.2	3231	4	US-09-134-001C-1200
27	32.6	2.2	5224	4	US-09-033-428-2

28	32.6	2.2	5224	4	US-09-033-556-5	Sequence 5, Appl
29	32.6	2.2	9751	4	US-09-238-303-7	Sequence 7, Appl
30	32.4	2.2	38844	4	US-09-734-675-3	Sequence 3, Appl
31	32.2	2.1	969	4	US-09-712-016-46	Sequence 46, Appl
32	32.2	2.1	1050	4	US-09-134-001C-2047	Sequence 2047, Ap
33	32.2	2.1	3319	3	US-08-335-844A-15	Sequence 15, Appl
34	32.2	2.1	5687	2	US-08-380-403A-3	Sequence 3, Appl
35	32.2	2.1	5687	2	US-08-895-628-3	Sequence 3, Appl
36	32.2	2.1	5687	4	US-08-895-610D-3	Sequence 3, Appl
37	32.2	2.1	51259	3	US-08-781-891-209	Sequence 209, App
38	32.2	2.1	162450	4	US-09-345-882-1	Sequence 1, Appl
39	32	2.1	1298	3	US-08-948-705-3	Sequence 3, Appl
40	32	2.1	72604	4	US-09-268-992-7	Sequence 7, Appl
41	32	2.1	72604	4	US-09-657-474-7	Sequence 7, Appl
42	31.8	2.1	289	4	US-09-007-005-17	Sequence 17, Appl
43	31.8	2.1	289	4	US-09-244-796-17	Sequence 17, Appl
44	31.8	2.1	1242	2	US-08-909-965C-5	Sequence 5, Appl
45	31.8	2.1	3144	4	US-08-961-527-224	Sequence 224, App

ALIGNMENTS

RESULT 1
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEITLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSER: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22113-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)883-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-F15
US-08-232-463-14
Query Match 4.9%; Score 74; DB 1; Length 7218;

Best Local Similarity 3.6%; Pred. No. 3.5e-13;
Matches 14; Conservative 236; Mismatches 136; Indels 0; Gaps 0;

```
OY 399 TCTGTCTCAATGCTCCCTACACTTCTCTCATTAATCCACATATCTAAAGCTATAGA 458
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1074 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1133
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 459 GATAATTTCATGCACTAGCTTACAGACATACCTATCTAGTCTACCTGTAACT 518
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1134 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1193
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 519 GGTATCTTCACATCTCGAATATTCATTTCTATGCTCAAGCTTGAGAG 578
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1194 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1253
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 579 TCTAGCTTATGATCTCCTTTCTCTCAGGTCTCTGCTCCACACACATTCAC 638
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1254 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1313
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 639 TTTCTGAATTTCTTGAACATTAACAAATCTCTCCATGGTGTGCTCCCTAC 698
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1314 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1373
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 699 TTTATGCTTCAGAGATCTACTCTGCCCATCTCATCTGCTTGGTCAATCA 758
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1374 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1433
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 759 AATCTCAATATGACCATTTCTTAA 784
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1434 YGTACCAATTTCTTCTATCTTAA 1459
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 2
US-09-810-347-3

```
; Sequence 3, Application US/09810347
; Patent No. 6461847
; GENERAL INFORMATION:
; APPLICANT: YE, JANE et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; FILE REFERENCE: CL001169
; CURRENT APPLICATION NUMBER: US/09/810.347
; CURRENT FILING DATE: 2001-03-19
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 42571
; TYPE: DNA
; ORGANISM: Human
US-09-810-347-3
```

Query Match 2.7%; Score 40.8; DB 4; Length 42571;
Best Local Similarity 51.7%; Pred. No. 0.057;
Matches 123; Conservative 0; Mismatches 107; Indels 8; Gaps 1;

```
OY 497 CTACTGTCTTACCTGTACTGTATCTGATGACATCTCGAATATTTCAATTCTCT 556
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 12370 CTTCAGGCTATTTGCTAAATCTGCTTCCATTTCAAGAAGATTTCCATATTTCT 12429
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 557 ATTGCTGCAAGTCTTGAAGAGTCTAGTCTTATGATCTCTTTCTCTCAGGTCTCT 616
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 12430 CTTCATGCCAGCAAAATCAATGTGTTCTCATCTACTGTTCTTCCACCTCCA 12489
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 617 GGTCTCCACACACATTCACACTCTTGA-----ATATCTTGAACATPAACAAAT 668
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 12490 GAGCATTCACATGCTACTGCTGATTCGCCCAATGATATCTACCATCCGACAT 12549
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 669 CTCTCATGGGTTGTTCCCTTACCCCAATTCATGCTCAGATCTACTTACTCTGCC 726
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 12550 TACTTCTCTCTTTTTCACATCTGCTCCTCATCTCTCATTTTCAAGACTCTTAA 12607
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 3
US-09-735-934A-3

```
; Sequence 3, Application US/09735934A
; Patent No. 6372468
; GENERAL INFORMATION:
; APPLICANT: LI, JIAYIN et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: CL000851
; CURRENT APPLICATION NUMBER: US/09/735.934A
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 43950
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-735-934A-3
```

Query Match 2.4%; Score 35.6; DB 4; Length 43950;
Best Local Similarity 58.5%; Pred. No. 2.7;
Matches 62; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

```
OY 1126 TGTTTGAGAGCCGCAAAATCTGACACCAATGTTTCAATGTTTAGCTGATGACACCTGT 1185
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 8258 TGGGTGGAAGGTGAAAAGAAAACAAACCTTGAGCTGGGACAGTGGTTCACACTGT 8317
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1186 AATCTGTACTTAGAGCAGATGCGAGGGGAGCTATGATTTAA 1231
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 8318 AACCACCACTTTGAGGCTGAGCGAGCGGATCATGAGTCAAG 8363
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 4
US-08-599-252-95/C

```
; Sequence 95, Application US/08599252
; Patent No. 5705343
; GENERAL INFORMATION:
; APPLICANT: DRAVINA, DENNIS T.
; APPLICANT: FEDER, JOHN N.
; APPLICANT: GIMKE, ANDREAS
; APPLICANT: KIMMEL, BRUCE E.
; APPLICANT: THOMAS, WINSTON J.
; APPLICANT: WOLFE, ROGER K.
; TITLE OF INVENTION: METHOD TO DIAGNOSE HEREDITARY
; TITLE OF INVENTION: HEMOCHROMATOSIS
; NUMBER OF SEQUENCES: 124
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & ROESTER
; STREET: 2000 Pennsylvania Ave. N.W., Suite 5500
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1888
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/599.252
; FILING DATE: 09-FEB-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 9053-0001.21
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 95:
; SEQUENCE CHARACTERISTICS:
```

LENGTH: 1166 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-599-252-95

Query Match 2.4%; Score 35.4; DB 1; Length 1166;
Best Local Similarity 73.8%; Pred. No. 0.31;
Matches 45; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

OY 1081 TAACCAACTTAATAGATTAAACGACGAGTTATCTCAGATGTTTGAGACGCCA 1140
DB 157 TACCACAACTTAATAGCTTAACACACAAATTATATCTCAGCTGTGATGTCA 98

OY 1141 G 1141
DB 97 G 97

RESULT 5

PCT-US96-06352-95/C
Sequence 95, Application PC/TUS9606352

GENERAL INFORMATION:
APPLICANT: DRAVNA, DENNIS T.
APPLICANT: FEDER, JOHN N.
APPLICANT: GRIKKE, ANDREAS
APPLICANT: KIMMEL, BRUCE E.
APPLICANT: THOMAS, WINSTON J.
APPLICANT: WOLFF, ROGER K.
TITLE OF INVENTION: METHOD TO DIAGNOSE HEREDITARY
TITLE OF INVENTION: HEMOCHROMATOSIS
NUMBER OF SEQUENCES: 124
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 Pennsylvania Ave. N.W., Suite 5500
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20006-1888

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/06352
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/599,252
FILING DATE: 09-FEB-1996
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 9053-0001.21
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0763
TELEX: 90-4030
INFORMATION FOR SEQ ID NO: 95:
SEQUENCE CHARACTERISTICS:
LENGTH: 1166 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
PCT-US96-06352-95

Query Match 2.4%; Score 35.4; DB 5; Length 1166;
Best Local Similarity 73.8%; Pred. No. 0.31;
Matches 45; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

OY 1081 TAACCAACTTAATAGATTAAACGACGAGTTATCTCAGATGTTTGAGACGCCA 1140
DB 157 TACCACAACTTAATAGCTTAACACACAAATTATATCTCAGCTGTGATGTCA 98

DB 157 TACCACAACTTAATAGCTTAACACACAAATTATATCTCAGCTGTGATGTCA 98
OY 1141 G 1141
DB 97 G 97

RESULT 6

PCT-US96-06583-95/C
Sequence 95, Application PC/TUS9606583

GENERAL INFORMATION:
APPLICANT: DRAVNA, DENNIS T.
APPLICANT: FEDER, JOHN N.
APPLICANT: GRIKKE, ANDREAS
APPLICANT: KIMMEL, BRUCE E.
APPLICANT: THOMAS, WINSTON J.
APPLICANT: WOLFF, ROGER K.
TITLE OF INVENTION: METHOD TO DIAGNOSE HEREDITARY
TITLE OF INVENTION: HEMOCHROMATOSIS
NUMBER OF SEQUENCES: 124
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 Pennsylvania Ave. N.W., Suite 5500
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20006-1888

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/06583
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/599,252
FILING DATE: 09-FEB-1996
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 9053-0001.21
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0763
TELEX: 90-4030
INFORMATION FOR SEQ ID NO: 95:
SEQUENCE CHARACTERISTICS:
LENGTH: 1166 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
PCT-US96-06583-95

Query Match 2.4%; Score 35.4; DB 5; Length 1166;
Best Local Similarity 73.8%; Pred. No. 0.31;
Matches 45; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

OY 1081 TAACCAACTTAATAGATTAAACGACGAGTTATCTCAGATGTTTGAGACGCCA 1140
DB 157 TACCACAACTTAATAGCTTAACACACAAATTATATCTCAGCTGTGATGTCA 98
OY 1141 G 1141
DB 97 G 97

RESULT 7

US-08-135-511-32
Sequence 32, Application US/08135511
Patent No. 5558999
GENERAL INFORMATION:

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,852
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/361,458
FILING DATE: 21-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/135,511
FILING DATE: 13-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/135,488
FILING DATE: 13-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/135,510
FILING DATE: 13-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: SANDERCOCK, Colin G.
REGISTRATION NUMBER: 31,298
REFERENCE/DOCKET NUMBER: 18748/220 HOCCE
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 5537 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-483-852-9

Query Match	2.48	Score 35.4	DB 1	Length 5537
Best Local Similarity	55.28	Pred. No. 0.85		
Matches 69	Conservative 0	Mismatches 56	Indels 0	Gaps 0
QY 742	TCTGCTTGTTCATTCGAAATCTCAAAATGTAGCAATTTCTTAAAGAGCTCTCCAGAGAAATA	801		
Db 1148	TCTCTCTTCTTGTCATTCAGATCTCAAAATGTCAAAATTTCAAGAGAGCCCATCTCTGATCATC	1207		
QY 802	ATATTGGAAGCATTTTGCTATTTCTATATCAAGTGAATCAATATGCTGCTCTGCCACCA	861		
Db 1208	ATATCTAAAGTTGTCCCTCATTCGCCCATAGCTTCTTATACCATGTGTTTATTTTTCATTA	1267		
QY 862	CCATG 866			
Db 1268	ACATG 1272			

RESULT 9
 US-08-361-458-4
 Sequence 4, Application US/08361458
 Patent No. 5663483
 GENERAL INFORMATION:
 APPLICANT: Chiang, John
 TITLE OF INVENTION: Genomic DNA of Human Cholesterol
 TITLE OF INVENTION: 7a-hydroxylase and Methods for using it
 NUMBER OF SEQUENCES: 6
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Foley & Lardner
 STREET: 3000 K Street, N.W., Suite 500
 CITY: Washington, D.C.
 COUNTRY: USA
 ZIP: 20007-5109
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/361,458
 FILING DATE:

TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 5537 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-187-453-32

Query Match 2.4%; Score 35.4; DB 1; Length 5537;
Best Local Similarity 55.2%; Pred. No. 0.85;
Matches 69; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 742 TCTGCTTGGTCATTCATCAATGTAGCCATTTCTAAAGGCTCTCCAGAGATA 801
DB 1148 TCTCTTTTGTTCATTCAGATCTCAATGTACAAATTCAGAGAGCCATCTGATCATC 1207
QY 802 ATATTGAAGCATTTTGTCTATTCATCAAGTATCATCAATGTCTGCTGCGACCA 861
DB 1208 ATATCTAAGTGTCTCATTCCTCCCATAGCTTCTATACAGTGTATTTTTCATA 1267
QY 862 CCATG 866
DB 1268 ACATG 1272

RESULT 12

US-08-562-985A-5
Sequence 5, Application US/08562985A
Patent No. 5821057
GENERAL INFORMATION:
APPLICANT: CHIANG, John Y.L.
TITLE OF INVENTION: AN ASSAY FOR AGENTS THAT AFFECT
TITLE OF INVENTION: CHOLESTEROL 7ALPHA-HYDROXYLASE EXPRESSION AND A
TITLE OF INVENTION: CHARACTERIZATION OF ITS REGULATORY ELEMENTS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/562,985A
FILING DATE: 27-NOV-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SANDERCOCK, Colin G.
REGISTRATION NUMBER: 31,298
REFERENCE/DOCKET NUMBER: 18748/213/HOCE
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 5537 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: join(2300..2380, 3930..4169, 5211..5537)

US-08-562-985A-5

Query Match 2.4%; Score 35.4; DB 1; Length 5537;
Best Local Similarity 55.2%; Pred. No. 0.85;
Matches 69; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 742 TCTGCTTGGTCATTCATCAATGTAGCCATTTCTAAAGGCTCTCCAGAGATA 801
DB 1148 TCTCTTTTGTTCATTCAGATCTCAATGTACAAATTCAGAGAGCCATCTGATCATC 1207
QY 802 ATATTGAAGCATTTTGTCTATTCATCAAGTATCATCAATGTCTGCTGCGACCA 861
DB 1208 ATATCTAAGTGTCTCATTCCTCCCATAGCTTCTATACAGTGTATTTTTCATA 1267
QY 862 CCATG 866
DB 1268 ACATG 1272

RESULT 13

US-08-477-952-9
Sequence 9, Application US/08477952
Patent No. 5851780
GENERAL INFORMATION:
APPLICANT: CHIANG, John Young Ling
TITLE OF INVENTION: Genomic DNA of Human Cholesterol
TITLE OF INVENTION: 7a-Hydroxylase and Methods of Using It
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,952
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/361,458
FILING DATE: 21-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/135,511
FILING DATE: 13-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/135,488
FILING DATE: 13-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/135,510
FILING DATE: 13-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: SANDERCOCK, Colin G.
REGISTRATION NUMBER: 31,298
REFERENCE/DOCKET NUMBER: 18748/221 HOCE
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 5537 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-477-952-9

Query Match

2.4%; Score 35.4; DB 2; Length 5537;

Best Local Similarity 55.2%; Pred. No. 0.85;
Matches 69; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 742 TCTGCTTGGTCATTCATAATCTCAATGTACGCAATTTCTAAAGGCTCTCCAGAGATA 801
DB 1148 TCTCTCTGTCTATGATCTCAATGTCAATTTTCAGAGAGCCCATCTCTGATCATC 1207

QY 802 ATATTGAAAGCATTTTGTCTATTCATCATGATCATCATGCTCTGCTGCCACCA 861

DB 1208 ATATCTAAAGTGTCTCATTCCTCCCATAGCTTCTCATACATGTTTATTTTTCARA 1267

QY 862 CCATG 866

DB 1268 ACATG 1272

RESULT 14

US-09-799-345-3/c

Sequence 3, Application US/09799345

Patent No. 6323016

GENERAL INFORMATION:

APPLICANT: YE, Jane et al.

TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC

TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES

TITLE OF INVENTION: THEREOF

FILE REFERENCE: CLO01156

CURRENT APPLICATION NUMBER: US/09/799,345

CURRENT FILING DATE: 2001-03-06

NUMBER OF SEQ ID NOS: 6

SOFTWARE: FASTSEQ for Windows Version 4.0

SEQ ID NO 3

LENGTH: 8758

TYPE: DNA

ORGANISM: Human

US-09-799-345-3

Query Match 2.4%; Score 35.4; DB 4; Length 8758;
Best Local Similarity 63.5%; Pred. No. 1.1;
Matches 54; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 1081 TAACACAACTTAATAGATTAACACGACGATTTATTCACATGTTTGGAGGCCA 1140

DB 4033 TACCACAAATTTAGTGGCTTAAACACACATTAATTTCTGTACAGTTGGAGGTCA 3974

QY 1141 GAAATCTGACACCGATTTCAATGTT 1165

DB 3973 GAAGCCCAAGCCACTTCACTGCT 3949

RESULT 15

US-09-442-143A-10

Sequence 10, Application US/09442143A

Patent No. 6403089

GENERAL INFORMATION:

APPLICANT: Levy, Gary

TITLE OF INVENTION: Methods of Modulating Immune Coagulation

FILE REFERENCE: 9579-14

CURRENT APPLICATION NUMBER: US/09/442,143A

CURRENT FILING DATE: 1999-11-15

PRIOR APPLICATION NUMBER: US 60/046,537

PRIOR FILING DATE: 1997-05-17

PRIOR APPLICATION NUMBER: US 60/061,684

PRIOR FILING DATE: 1997-10-10

NUMBER OF SEQ ID NOS: 53

SOFTWARE: PatentIn version 3.1

SEQ ID NO 10

LENGTH: 1339

TYPE: DNA

ORGANISM: Murine

US-09-442-143A-10

Query Match 2.2%; Score 33.6; DB 4; Length 1339;

Best Local Similarity 70.3%; Pred. No. 1.3;
Matches 45; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 1172 TGATGCACCTGTAATCTGTGTAATAGAGGAGATGAGGAGCTGATTAAG 1231

DB 389 TGGTACACACCTGTGATCTGTGTTAGAGGAGAGGAGAGATTCAGAGTTCAAG 448

QY 1232 CCCA 1235

DB 449 GCCA 452

Search completed: February 17, 2003, 16:11:05
Job time: 125.486 secs

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

(without alignments)
11653.717 Million cell updates/sec

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Maximum DB seq length: 200000000000

Post-processing:	Minimum Match	08
	Maximum Match	1000

Listing first 45 summaries

Database :

```

1:  GenMem1.*
2:  gb_ba.*
3:  gb_htg.*
4:  gb_in.*
5:  gb_om.*
6:  gb_ov.*
7:  gb_pat.*
8:  gb_ph.*
9:  gb_pl.*
10: gb_pr.*
11: gb_ro.*
12: gb_sis.*
13: gb_sy.*
14: gb_v1.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sis.*
28: em_sts.*
29: em_un.*
30: em_v1.*
31: em_hg_hum.*
32: em_hg_inv.*
33: em_hg_other.*
34: em_hg_pln.*
35: em_hg_pln.*
36: em_hg_rod.*
37: em_hg_mam.*
38: em_hg_vrt.*
39: em_hggo_hum.*
40: em_hggo_mus.*
41: em_hggo_other.*

```

Pred. No. is the number of results predicted by chance to have a

Result No.	Score	Query Match	Length	DB	ID	Description
1	1494.4	99.6	219386	2	AC125147	AC125147 Mus muscu
2	743	49.5	187332	2	AC110401	AC110401 Rattus no
3	732.6	48.8	165503	2	AC111518	AC111518 Rattus no
4	98	6.5	168887	9	AC106796	AC106796 Homo sapi
5	74	4.9	7218	6	166494	166494 Sequence 14
6	57.2	3.8	162548	9	AC015831	AC015831 Homo sapi
7	53	3.5	128484	9	AC025270	AC025270 Homo sapi
8	53	3.5	162495	9	AC009996	AC009996 Homo sapi
9	52.6	3.5	98151	9	AL359543	AL359543 Human DNA
10	51.4	3.4	178985	9	AC015524	AC015524 Homo sapi
11	51.4	3.4	201156	2	AC021757	AC021757 Homo sapi
12	50.4	3.4	167746	9	AC011120	AC011120 Homo sapi
13	50.4	3.4	172884	9	AC113567	AC113567 Canis fam
14	50.4	3.4	193209	2	AC117938	AC117938 Canis fam
15	50	3.3	134019	9	AC003667	AC003667 Homo sapi
16	50	3.3	160176	2	AC025239	AC025239 Homo sapi
17	49.6	3.3	100000	9	AC117455	Continuation (2 of
18	49.6	3.3	328883	2	AC117455	AC117455 Homo sapi
19	49	3.3	154429	2	AC090381	AC090381 Homo sapi
20	49	3.3	156467	2	AC083771	AC083771 Homo sapi
21	49	3.3	156576	2	AP001371	AP001371 Homo sapi
22	49	3.3	164554	9	AC084811	AC084811 Homo sapi
23	49	3.3	167292	9	AC016205	AC016205 Homo sapi
24	48.4	3.2	164813	9	AC067943	AC067943 Homo sapi
25	48	3.2	107226	9	AC010618	AC010618 Homo sapi
26	48	3.2	109882	2	AC131243	AC131243 Homo sapi
27	48	3.2	129766	9	AC104942	AC104942 Homo sapi
28	48	3.2	166797	2	AC016718	AC016718 Homo sapi
29	47.8	3.2	117755	9	AC092686	AC092686 Homo sapi
30	47.8	3.2	160628	2	AC067756	AC067756 Homo sapi
31	47.8	3.2	168463	9	AC021463	AC021463 Homo sapi
32	47.8	3.2	170999	9	AC093267	AC093267 Homo sapi
33	47.8	3.2	194691	2	AC068221	AC068221 Homo sapi
34	47.6	3.2	94924	9	AL355997	AL355997 Human DNA
35	47.6	3.2	132200	2	AL161619	AL161619 Homo sapi
36	47.6	3.2	178212	2	AC013326	AC013326 Homo sapi
37	47.6	3.2	200333	2	AC073279	AC073279 Homo sapi
38	47.4	3.2	100418	9	HS102662	HS102662 Human DNA
39	47.4	3.2	168751	2	AC066962	AC066962 Homo sapi
40	47.4	3.2	170883	9	AC067805	AC067805 Homo sapi
41	47.4	3.2	188736	9	CNS00DM8T	AL079305 Human chr
42	47.4	3.2	196933	9	CNS00DM8T	AL079305 Human chr
43	47.2	3.1	61246	2	AL350032	Continuation (5 of
44	47.2	3.1	156813	9	HS1631M13	AL17333 Human DNA
45	47.2	3.1	161273	9	HS769N13	AL035427 Human DNA

ALIGNMENTS

RESULT	1
AC125147	
LOCUS	AC125147
DEFINITION	AC125147 219386 bp DNA linear HTG 22-JUN-2007
ACCSSION	Mus musculus chromosome UKK clone RP24-350A15, WORKING DRAFT
VERSION	AC125147
KEYWORDS	SEQUENCE, 8 unordered pieces.
SOURCE	AC125147.2 GI:21539180
ORGANISM	HTG; HTGS_PHASE1; HTGS_DRAFT. house mouse. <i>Mus musculus</i> Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 219386)
AUTHORS	McPherson,J.D. and Waterston,R.H.
TITLE	The sequence of Mus musculus clone

JOURNAL
REFERENCE 2 (bases 1 to 219386)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 219386)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (22-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
On Jun 22, 2002 this sequence version replaced gi:21490647.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
Project Information
Center project name: M_BB0350A15

----- Summary Statistics -----
Sequencing vector: MJ3: 0%
Chemistry: Dye-Primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 216121 bases at least Q40
Consensus quality: 216506 bases at least Q30
Consensus quality: 216719 bases at least Q20
Insert size: 171000; agarose-fp
Insert size: 217980; sum-of-contigs
Quality coverage: 14.92 in Q20 bases; sum-of-contigs
Quality coverage: 11.97 in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 105: contig of 105 bp in length
* 106 205: gap of unknown length
* 206 806: contig of 601 bp in length
* 807 906: gap of unknown length
* 907 5289: contig of 4363 bp in length
* 5270 5369: gap of unknown length
* 5370 16393: contig of 11024 bp in length
* 16394 34651: gap of unknown length
* 34652 34751: contig of 18158 bp in length
* 34752 80845: gap of unknown length
* 80846 80945: gap of unknown length
* 80946 138242: contig of 57297 bp in length
* 138243 219386: gap of unknown length
* 219387 219386: contig of 81044 bp in length.

FEATURES
SOURCE
1. 219386
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP24-350A15"
misc_feature
1..105
/note="assembly_name:Contig12"
misc_feature
206..806
/note="assembly_name:Contig14"
misc_feature
907..5269
/note="assembly_name:Contig19"
misc_feature
5370..16393
/note="assembly_name:Contig20"
misc_feature
16494..34651

misc_feature
34752..80845
/note="assembly_name:Contig21"
misc_feature
80946..138242
/note="assembly_name:Contig23"
misc_feature
138343..219386
/note="assembly_name:Contig24"
BASE COUNT 60953 a 47433 c 47470 g 62799 t 731 others
ORIGIN

Query Match 99.6%; Score 1494.4; DB 2; Length 219386;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1495; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

5 GGGCCCTGGGAGTGTGGCTAGTCTTGGCAATGACCTGGAGACAGCTTGGCCATA 64
Db 118700 GGGCCCTGGGAGTGTGGCTAGTCTTGGCAATGACCTGGAGACAGCTTGGCCATA 118759

65 TGAGATCCAGTGACAAAGCTCATCTAGATGCTGTGATACCAATTAAGTGAACCATCATTA 124
Db 118760 TGAGATCCAGTGACAAAGCTCATCTAGATGCTGTGATACCAATTAAGTGAACCATCATTA 118819

125 TGCAATCAGCCGGAAGTATCTCTGTGGCTTGTCTTCTTACTGTAATGATTAAC 184
Db 118820 TGCAATCAGCCGGAAGTATCTCTGTGGCTTGTCTTCTTACTGTAATGATTAAC 118879

185 CTATATGATTTTACCATTTTCCCTCCATGGAGACATCCTCTCTTATGTGACCC 244
Db 118880 CTATATGATTTTACCATTTTCCCTCCATGGAGACATCCTCTCTTATGTGACCC 118939

245 TACTTATGTCCTATGTCAGTCCAGCTGCTTCTTGTATGAGAGACCAATCTGTTCTTCTA 304
Db 118940 TACTTATGTCCTATGTCAGTCCAGCTGCTTCTTGTATGAGAGACCAATCTGTTCTTCTA 118999

305 TGTCATCTGCTCACTTTTCCAGCTGACTCCACCATCTGTTACTATTCAGAGTCACT 364
Db 119000 TGTCATCTGCTCACTTTTCCAGCTGACTCCACCATCTGTTACTATTCAGAGTCACT 119059

365 CACAGTTCTTGAGAGCAGAGACATCAGACATGATGTCCTCAATGCTCCCTACACT 424
Db 119060 CACAGTTCTTGAGAGCAGAGACATCAGACATGATGTCCTCAATGCTCCCTACACT 119119

425 TTCTCCATTAATCCATATCTTAAGCTATAGAGATTAATTCATGACATTAAGCTTCA 484
Db 119120 TTCTCCATTAATCCATATCTTAAGCTATAGAGATTAATTCATGACATTAAGCTTCA 119179

485 GTACTATGCTATGCTACTGCTCTTACCTGTAACCTGTAATCTTGAATATTT 544
Db 119180 GTACTATGCTATGCTACTGCTCTTACCTGTAACCTGTAATCTTGAATATTT 119239

545 TTCAATTTCTATTTGTCGCAAGCTGAGAGCTACTTATGAGATCTCTTTCTC 604
Db 119240 TTCAATTTCTATTTGTCGCAAGCTGAGAGCTACTTATGAGATCTCTTTCTC 119299

605 CTCAGTCTCCTGCTCCACACACATTCACATCTTGTGAATATCTTTGAACATTA 664
Db 119300 CTCAGTCTCCTGCTCCACACACATTCACATCTTGTGAATATCTTTGAACATTA 119359

665 AATTCTTCATGGGTTTGTCCCTTAACCAATTCATGCTTCAGATCTTACTCTG 724
Db 119360 AATTCTTCATGGGTTTGTCCCTTAACCAATTCATGCTTCAGATCTTACTCTG 119419

725 CCCCATCTCATCTATCTCTCTTTGGTCAATTCGAATGAGCATTTCTTAATA 784
Db 119420 CCCCATCTCATCTATCTCTCTTTGGTCAATTCGAATGAGCATTTCTTAATA 119479

785 GGCTCTCCAGAGAAATATATTTGAAGCATTTTCTATCTATCAAGATCATCAAT 844
Db 119480 GGCTCTCCAGAGAAATATATTTGAAGCATTTTCTATCTATCAAGATCATCAAT 119539

845 GTCTGCTCTGACCAACCATGACCATCCCATGAATACAGACACTGCTTCTTACTGTT 904
Db 119540 GTCTGCTCTGACCAACCATGACCATCCCATGAATACAGACACTGCTTCTTACTGTT 119599


```

* 6552: gap of unknown length
* 6552: contig of 1461 bp in length
* 8013: gap of unknown length
* 8112: gap of unknown length
* 9240: contig of 1128 bp in length
* 9340: gap of unknown length
* 9341: gap of unknown length
* 10703: contig of 1363 bp in length
* 10803: gap of unknown length
* 10804: contig of 2074 bp in length
* 12878: gap of unknown length
* 12978: gap of unknown length
* 12978: contig of 1321 bp in length
* 14298: gap of unknown length
* 14398: gap of unknown length
* 15844: contig of 1446 bp in length
* 15845: gap of unknown length
* 15945: contig of 1643 bp in length
* 17587: gap of unknown length
* 17688: gap of unknown length
* 17688: contig of 1046 bp in length
* 18734: gap of unknown length
* 18834: gap of unknown length
* 20948: contig of 2115 bp in length
* 21048: gap of unknown length
* 23317: contig of 2269 bp in length
* 23417: gap of unknown length
* 23418: contig of 2012 bp in length
* 25430: gap of unknown length
* 25530: contig of 1528 bp in length
* 27058: gap of unknown length
* 27158: contig of 1663 bp in length
* 28820: gap of unknown length
* 28821: contig of 1338 bp in length
* 30258: gap of unknown length
* 30359: contig of 2216 bp in length
* 32574: gap of unknown length
* 32575: contig of 1640 bp in length
* 34315: gap of unknown length
* 34415: gap of unknown length
* 36024: contig of 1610 bp in length
* 36125: gap of unknown length
* 38547: contig of 2423 bp in length
* 38548: gap of unknown length
* 41952: contig of 3305 bp in length
* 42052: gap of unknown length
* 42053: contig of 2116 bp in length
* 44169: gap of unknown length
* 44268: contig of 2760 bp in length
* 47029: gap of unknown length
* 47128: gap of unknown length
* 47129: contig of 1920 bp in length
* 49048: gap of unknown length
* 49148: gap of unknown length
* 50748: contig of 1600 bp in length
* 50848: gap of unknown length
* 53484: contig of 2636 bp in length
* 53485: gap of unknown length
* 55862: contig of 2278 bp in length
* 55863: gap of unknown length
* 57817: contig of 1855 bp in length
* 57917: gap of unknown length
* 60973: contig of 3056 bp in length
* 61073: gap of unknown length
* 64269: contig of 3196 bp in length
* 64270: gap of unknown length
* 64370: gap of unknown length
* 67983: contig of 3614 bp in length
* 68083: gap of unknown length
* 71383: contig of 3300 bp in length
* 71384: gap of unknown length
* 75737: contig of 4254 bp in length
* 75838: gap of unknown length
* 75838: contig of 3683 bp in length
* 79521: gap of unknown length
* 79620: gap of unknown length
* 83421: contig of 3801 bp in length
* 83521: gap of unknown length
* 83522: contig of 3993 bp in length
* 87515: gap of unknown length
* 87515: contig of 4497 bp in length
* 92112: gap of unknown length
* 92112: contig of 4323 bp in length
* 96535: gap of unknown length

```

```

* 96635 101680: contig of 5046 bp in length
* 101681 101780: gap of unknown length
* 101781 107239: contig of 5459 bp in length
* 107240 107339: gap of unknown length
* 107340 113255: contig of 5916 bp in length
* 113256 113355: gap of unknown length
* 113356 118196: contig of 4841 bp in length
* 118197 118296: gap of unknown length
* 118297 121989: contig of 3693 bp in length
* 121990 122089: gap of unknown length
* 122090 126651: contig of 4562 bp in length
* 126652 126751: gap of unknown length
* 126752 129765: contig of 3014 bp in length
* 129766 129865: gap of unknown length
* 129866 137350: contig of 7485 bp in length
* 137351 137450: gap of unknown length
* 137451 142649: contig of 5199 bp in length
* 142650 142749: gap of unknown length
* 142750 148513: contig of 5764 bp in length
* 148514 148613: gap of unknown length
* 148614 155533: contig of 6920 bp in length
* 155534 155633: gap of unknown length
* 155634 162354: contig of 6721 bp in length
* 162355 162454: gap of unknown length
* 162455 170099: contig of 7645 bp in length

Query Match      49.5%: Score 743; DB 2; Length 187332;
Best Local Similarity 76.1%: Pred. No. 3.6e-184;
Matches 1140; Conservative 0; Mismatches 205; Indels 154; Gaps 12;

QY 5 GGGCCCTCGGAGCTTTGGCTAAGCTTGCAGTAAGCTGTGATGAGAGTTGGCCATTA 64
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 172697 GGGCCCTCATGACTTTGGCT- -GCTTGCAGTAGATGTATAGCCAGCTTGCACCATTA 172754

QY 65 TGAATCCAGTGAACAAGCTCTCTAGATGCTGCATACCAATAGTAGCCATCATTA 124
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 172755 AGAATCCAAATGCCAACCTGTGCTGTGATGATAGAGCTAGATGATGACCATCA 172814

QY 125 TGAATCAGGCGGAGCATCCTCTGTGGCTTTGTCTTCTTACTGATTAATTTGATAC 184
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 172815 TGAAGTTAGGCCAGAAATCCTCTGTGACCTTTCTTACTATTAATTTGACGAC 172874

QY 185 CTTATGATTTTCCCTTTCCCTTCATGAGCATCACTCCCTTCCCTTGTGACCC 244
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 172875 CTTATGATTTTACCA- -TTCTCTCCATGAGCATTCACGCTCTTTCACAGACGCC 172933

QY 245 TACTTATGCTCTATGATGATCAGCTGCTCTTGTATGAGAGCCATCCTGTTCTTCTA 304
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 172934 CACCTCT- -TCCFATGATGATCAGCTG- -TCCCTTGTATGAGAGCCATCCTGATCTTCTA 172991

QY 305 TGTGACTGTGCTCACTTCTTCACGTAAGTCACTCAATCTGTCTACATTTGACAGTCACT 364
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 172992 T- - - - -GTGACTCCACTTAAGCTCTGTATGATGAGAGTCACTTCACT 173028

QY 365 CACAGTTTCTTGAGAGAGAGACTCAGAACTGATCTGCTCAATGTCCTCCTACACT 424
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 173029 CATAGTTTCTTTAGAGAGAGA- - - - - 173051

QY 425 TTCTCTCATTAATCCACATATCTTAAGCTATAGAGATTAATTCATGACTATAGCTTCA 484
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 173052 - - - - - 173051

QY 485 GTRACTATCGATTTACTGTCTTACCCCTGTAACGTGATCTTTCATGACATCTGCAATATT 544
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 173052 - - - - -TGCTCTCTACCTCTGTAACGTGATCTTTCATGACATTTAGAGTATT 173098

QY 545 TCCAATTTCTCTATGCTGGAAGTCTTGAAGTCTAGTCTTATGATCTCTTCTC 604
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 173099 TCCAATTTCTCTATGCTGGAAGTCTTGAAGTCTAGTCTTATGATCTCTTCTC 173158

QY 605 CTCAGGCTCTGCTGCTGTCACACACCATTCACACTCTTGAATATTCTTGAACATTAACA 664
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 173159 CTCAGGCTCTGCTGCTGTCACACCATTCACACTCTTGAATATTCTTGAACATTAACA 173218

```

QY	665	AAATCTCTCCATGAGGGTTTGTTCOCCTTACCCAAATTCATGCGCTGAGTAATCTACTC--	722
Db	173219		
QY	723	TGCCCCATCTTCATCTATCTCTGCTTGGCTATTCAAATCTGCAAAATGAGCAATCTCTAA	782
Db	173279		
QY	783	AAGGCTCTCCAAAGAGAAATATATTGAAAGCATTTTGGTATTCATCAAGTGATCTACA	842
Db	173339		
QY	843	ATGTCTGCTCCTGACCAACCCATGAGACATCCCATGAAATACAGACACCGCTTCTAGAG	902
Db	173399		
QY	903	TTTGCTGATATGTTCTGTGTGTGATACATTTGATATGAATGCTATATAAATCTCTGTGA	962
Db	173459		
QY	963	GCAAAATTGAATATATCAGATAGACACCCCTCTCTGTGAGAGCAATGATCTCATGTTATCCC	1022
Db	173512		
QY	1023	AAAGCATGAGGTAAGAGACATTTATCCAGGTCCATGCTGGTTCCGATATGATTTGTTTCA	1082
Db	173572		
QY	1083	ACACAAATTAATTAATTAATAAAGACAGACGGATTTATCTCACATGTTTTGAGACGGCAGA	1142
Db	173632		
QY	1143	AATCGACACACAGTTTCAATGTTTAGCTGTATGACACACCTGTAAATCTTGCTGACTTACGA	1202
Db	173652		
QY	1203	GGCAGATGCA--GGGGGACTATGATTTAAAGCCATTTTAAAGCTGCTGGTGAGAACT	1260
Db	173748		
QY	1261	GTCCTGATTTTTTTTTCACATTTGGGCTAAAGTCAAGGATCAAGAGTTGGTGCAATCT	1320
Db	173807		
QY	1321	GGAAGAAACCTTTGGCTTGACACTTCCAGAGAGGGCGCAGATTCCTTGGTGTTGTT	1380
Db	173867		
QY	1381	GGTCTGGAATCACTGTGACCTTATGCTCCATCTCACAATTCCTCTGCAATTTATCTCT	1440
Db	173927		
QY	1441	AAGCAGCGGTGCTGTATATCCAACTTTAGAGGCCCATAGATCCCCCATTTTCTCTC	1499
Db	173987		
QY	1499	AAAGCATCTGTGGCTGTGTGTCCACCTTTAGGACCTCTCCACATCTCCCATCTGCTCTC	1559
Db	174045		

TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 165503)
AUTHORS	Worley, K. C.
TITLE	Direct Submission
JOURNAL	Submitted (19-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	3 (bases 1 to 165503)
AUTHORS	Worley, K. C.
TITLE	Direct Submission
JOURNAL	Submitted (23-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT	On Jul 12, 2002 this sequence version replaced gi:18701283.

TITLE

Albrooks,S.L., Amaraturunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Blinage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Burney,C., Burich,P., Burkett,C., Butrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chaves,D.,
Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Devila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
DeJaney,K.R., Delgado,O., Dem,A.L., Ding,Y., Dinh,H.J.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Dumbin,K.H.,
Eamhardt,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frenzt,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J., H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homsli,F., Howard,S., Huber,J., Huliyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Kovach,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Lounsged,H.,
Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., McWhiney,E., McLeod,M.P., Medora,M., Mel,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Mosier,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokkenko,S., Oguh,M., Okunnu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rivers,M., Rojas,A., Rojoubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scheier,S., Scott,G., Shen,H., Shoostlari,N., Sisson,I.,
Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamezisa,A., Tamezisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telirod,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalob,D., Yanson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Woreley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

Direct Submission

RESULT 3	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
AC111518	AC111518	165503 bp DNA							
		<i>Rattus norvegicus</i> clone CH230-142F15, *** SEQUENCING IN PROGRESS							
		***, 74 unordered pieces.							
	AC111518								
	AC111518.2	GI:21736068							
	HTG; HTGS-PHASE1.								
	<i>Rattus norvegicus</i> .								
	<i>Rattus norvegicus</i>								
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;								
	<i>Rattus</i> .								
	1 (bases 1 to 165503)								
	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,								

Alshrooks, S.L., Amaraltinge, H.C., Are, J.R., Ayele, M., Banks, T.,
Barbata, J., Benton, J., Blimace, K., Blankenbush, K., Bonini, D.,
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
Bunay, C., Butch, P., Burkett, C., Burrill, K.L., Byrd, N.C.,
Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chaver, D.,
Chen, G., Chen, R., Chen, Z., Chowdhury, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Davila, M.L., Davis, C., Day-Carroll, L., Dederich, D.A.,
Delaney, K.R., Delgado, O., Dem, A.L., Ding, Y., Dinh, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Dublin, K.J.,
Eamhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
Falls, T., Ferraruto, D., Flagg, N., Ford, J., Foster, P., Frentz, P.,
Gabriel, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorell, J.H., Guevara, W., Gunatane, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
Homs, I.F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
Kratoch, J., Kuesh, A., Landry, N., Deal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Lichtarge, O., Ilic, C., Liu, J., Liu, W., Louised, H.,
Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Meheswarai, M., Mapua, P., Martin, R., Martindale, A., Mattheis, E.,
Messer, E., Mawhine, E., Mcleod, M.P., Meador, M., Mel, G., Metzger, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
Moser, M., Neel, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokkenko, S., Ogul, M., Okunolu, G.,
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,
Scherer, S., Scott, G., Shen, H., Shostetari, N., Sisson, I.,
Sodegrien, E., Sonake, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Swatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tatney, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wleciyk, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y.F., Zhou, T., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.

Unpublished
Direct Submission
2 (phases 1 to 165503)
Worley, K.C.
Direct Submission
Submitted (19-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (phases 1 to 165503)
Worley, K.C.
Direct Submission
Submitted (23-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 12, 2002 this sequence version replaced gi:18701283.

Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: GMD
Center clone name: CH230-142P15
Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 85091 bases at least Q40
Consensus quality: 91433 bases at least Q30
Consensus quality: 97328 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This is a working draft sequence. It currently


```

QY 425 TTCTCTCATATATCCACATATCTAAAGCTATAGAGATTAATTCGACCTATAGCTTTCA 484
Db 154181 ----- 154180
QY 485 GTACTATGATATCTAGTCTCTACCTGTAACCTGATCTTACATACATCCGATATAT 544
Db 154181 -----TGCTCTCTACCTGTAACCTGTAACCTGATCTTACATACATCCGATATAT 154227
QY 545 TCCAAATTCCTATATGCTGCAAAAGCTTGAGAGTCTAGTCTTATGATCTCTCTCTC 604
Db 154228 TCCAAATTCCTATATGCTGCAAAAGCTTGAGAGTCTAGTCTTATGATCTCTCTCTC 154287
QY 605 CTGAGTCTCTCTGCTCTCCACACACCACTTACACATCTTGAATATCTTGAACATPACA 664
Db 154288 CTGAGTCTCTCTGCTCTCCACACACCACTTACACATCTTGAATATCTTGAAGATPACA 154347
QY 665 AATTCCTCATAGGGTGTGTTCCCTCTACCCCAATTCATGCTTACGATCTTACTCTC-- 722
Db 154348 AATTCCTCTGTGGGTTGTTCCCTTACCCCAAGCAGCATCTTACGATCTTACTCTC 154407
QY 723 TGCCCATCTTCTACTCATCTCTGCTTGTGCTATCAAAATCTCAAAATGATGCACTTCTTAA 782
Db 154408 CCCCCATTTTCACTCATCTCTAGTGTGTCTATTCATTCATTCATTCATTCATTCATTCAT 154467
QY 783 AAGCTCTCCAGAGAAATATATTTGAAAAGCATTTTGTCTATTCATTCATTCATTCAT 842
Db 154468 AAGCTCTCCAGAGAAATATATTTGAAAAGCATTTTGTCTATTCATTCATTCATTCAT 154527
QY 843 ATGTCTCTCTCTGCTCTACCATCATGACCATCTCCCATGATATACAGACATCCCTCTTAA 902
Db 154528 ATGTCTCTCTCTGCTCTACCATCATGACCATCTCCCATGATATACAGACATCCCTCTTAA 154587
QY 903 TTTCGTATATGTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 962
Db 154588 TTTCGTATGTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 154640
QY 963 GCATATGATATCATCATATAGACACCTCTCTCTGAGAGCATGATCTCATGTTATCC 1022
Db 154641 GCATATGATATCATCATATAGACACCTCTCTCTGAGAGCATGATCTCATGTTATCC 154700
QY 1023 AAGCATGAGATGATGAGATGATATCCAGGTCATGCTGTTCCGATGATGATGTTCTTA 1082
Db 154701 AAGCATGAGATGATGAGATGATATCCAGGTCATGCTGTTCCGATGATGATGTTCTTA 154760
QY 1083 ACACAAATTAATAGATTAATAACAGACAGATTTATCTCATATGTTTGAACAGCCAGA 1142
Db 154761 ACACAGATGATATAGATTAATAACAGACAGATTTATCTCATATGTTTGAACAGCCAGA 154820
QY 1143 AATCTGACACCATGTTTAAATGTTAGACTGTATGACACACCTGTATTTCTGTACTTGA 1202
Db 154821 AATCTGACACCATGTTTAAATGTTAGACTGTATGACACACCTGTATTTCTGTACTTGA 154876
QY 1203 GAGAGATGCA--GGGAGATGATGATTAAGCCATTTTAAAGCTGCTGAGGAGAGACCT 1260
Db 154877 GAGAGATGCA--GGGAGATGATGATTAAGCCATTTTAAAGCTGCTGAGGAGAGACCT 154935
QY 1261 GTCTGATTTTTTTTTCATATGAGCTTAAAGATCAAGATCATCAGGTTGCTGATCTT 1320
Db 154936 GTCTGATTTTTTTTTCATATGAGCTTAAAGATCAAGATCATCAGGTTGCTGATCTT 154994
QY 1321 GGAAGAAACCTTGTGCTGCAAGCTTCCAGAGGCGCCAGCATCTGCTGCTGTGTTT 1380
Db 154995 GGAAGAAACCTTGTGCTGCAAGCTTCCAGAGGCGCCAGCATCTGCTGCTGTGTTT 155054
QY 1381 GGTCCTGATCATCTGTGACCTTATGCTCATTCGACATTCCTGATCTTATTCCTCT 1440
Db 155055 GGTCCTGATCATCTGTGACCTTATGCTCATTCGACATTCCTGATCTTATTCCTCT 155114
QY 1441 AAGCACCAGGCTGCTGTATCAACATTTAGAGAGCCCATGATGATCCCATTTCTCTC 1499
Db 155115 AAGCACCAGGCTGCTGTATCAACATTTAGAGAGCCCATGATGATCCCATTTCTCTC 155173

```

```

RESULT 4
AC106796/c
LOCUS 168887 bp DNA linear PRI 06-AUG-2002
DEFINITION Homo sapiens chromosome 16 clone RP11-429K17, complete sequence.
AC106796
VERSION AC106796.2 GI:22122880
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 168887)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 168887)
AUTHORS DOE Joint Genome Institute.
TITLE DOE Joint Genome Institute.
JOURNAL Direct Submission
REFERENCE 3 (bases 1 to 168887)
AUTHORS DOE Joint Genome Institute.
TITLE DOE Joint Genome Institute.
JOURNAL Direct Submission
REFERENCE 4 (06-AUG-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Aug 6, 2002 this sequence version replaced gi:18139346.
COMMENT This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
Genbank flat file format but are available as part
of this entry's ASN.1 file.
-----
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
Genbank flat file format but are available as part
of this entry's ASN.1 file.
-----
FEATURES
source Location/Qualifiers
1..168887
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-429K17"
BASE COUNT 46572 a 37397 c 37923 g 46995 t
ORIGIN
Query Match 6.5%; Score 98; DB 9; Length 168887;
Best Local Similarity 51.9%; Pred. No. 5e-15;
Matches 336; Conservative 0; Mismatches 285; Indels 27; Gaps 4;
QY 555 CTATTCGTCGCAAGTCTTGAGAGTCTAGTCTTATGATCTCTTTCTCTAGCTCTC 614
Db 125148 CCAATGCTTCAACACCTTGAAAGGCTGCGCCACAGATCTCTCTCTCTCTCTCT 125089
QY 615 CTGCTCTCCACACACATTCATCTTGAATATCTTGAACATTAACAAATTCCTCC 674
Db 125088 TCACATTCAGAGCCGCCACACACATCTCTGCTCTATCTATGAAACCAATTCAC 125029
QY 675 ATGGGTTGTCTCCCTACCAACATTCATGCTTCAGGATCTACTCTGCCCCATCTTC 734
Db 125028 CTCTCATGTTCCCTCTGCGCTTAATGCACTCCCTTCAAAATTCATTTCCCGAGTT 124969
QY 735 ACTCATCTCTGCTTGTGTC-----ATTCAATCTCAATATGATGACCATTTCTAAAGG 786

```


[illegible]

LOCUS	166494	7218 bp	DNA	linear	PAT 28-DEC-1997
DEFINITION	Sequence 14 from patent US 5670367.				
ACCESSION	166494				
VERSION	166494.1	GI:2724471			
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	1 (bases 1 to 7218)				
AUTHORS	Donner, F., Scheifflinger, F. and Falkner, F. Gunter.				
TITLE	Recombinant fowlpox virus				
JOURNAL	Patent: US 5670367-A 14 23-SEP-1997;				
FEATURES	Location/Qualifiers				
Source	1..7218				
BASE COUNT	1944 a 1491 c 1486 g 1929 t	368 others			
ORIGIN					
Query Match	4.9%; Score 74; DB 6; Length 7218;				
Best Local Similarity	3.6%; Pred. No. 1,3e-08;				
Matches 14; Conservative 236; Mismatches 136; Indels 0; Gaps 0					
QY	399 TCTGCTCAATAGTCTCTCCCTCACTTTCTCCTCATATACCAATATCAAGCTATAGA	458			
Db	1074 YY	1133			
QY	459 GATATTTACGACGACTAATAGCTTTAGAGTATTCGATCTACTGCTGCTACCCCTGTA	518			
Db	1134 YY	1193			
QY	519 GGTATCTCATGACATCGCAATATTTCCATTTCTGCTGCGAAGCTGTAGAG	578			
Db	1194 YY	1253			
QY	579 TCTACTCTATAGACTCTCTTTCTCTCAGGCTCTGCTGCTCAGACACCATTCACAC	638			
Db	1254 YY	1313			
QY	639 TTCTGAAATATTCCTTGACATACAAATTCCTCGATGGGTTTGTCCCTTACCCAAA	698			

Dd	1314	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	1373
Oy	699	TTCATGCGTCCAGAACTACTACGTGCCCATCTTCACATCCTGGTCAATCA	758
Dd	1374	YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY	1433
Oy	759	AATCTCAAATGATGACCATTTCTPAAA	784
Dd	1434	YGTACCAAATCTCTCATCTCTTTA	1459

RESULT	6	
AC015831/c		
LOCUS	AC015831	
DEFINITION	Homo sapiens clone RP11-668G24, LOW-PASS	162548 bp DNA
ACCESSION	AC015831	linear
VERSION	AC015831.2	HTG 26-MAR-2000
KEYWORDS	HTG; HTGS-PHASED.	SEQUENCE SAMPLING.
SOURCE	Homo sapiens.	
ORGANISM	Homo sapiens	

REFERENCE	1 (bases 1 to 162548)
AUTHORS	Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE	Homo sapiens, clone RP11-688624
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 162548)
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.

Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lenocksky, J., Lieu, C., Locke, K., MacDonald, P., Marquis, N., McMan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Rile, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Tirelli, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W.-J., Zimmer, A., and Zody, M.

TITLE	Direct Submission
JOURNAL	Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT	On Mar 26, 2000 this sequence version replaced gi:5446837.

All repeats were identified using ReSmit, A.F.A. & Green, P. (1996-1997)

----- Genome Center
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information -----

Center project name: L4338
Center clone name: 688_G_24

* NOTE: This record contains 173 individual

- * sequencing reads that have not been assembled into contigs. Runs of N are used to separate the reads

- * and the order in which they appear is completely arbitrary. low-pass sequence sampling is useful for

- * identifying clones that may be gene-rich and allows
- * overlap relationships among clones to be deduced.

- * However, it should not be assumed that this clone
- * will be sequenced to completion. In the event that

- * the record is updated, the accession number will be preserved.

```

1      858: contig of 858 bp in length
*      859 958: gap of 100 bp
*

```

* 959 1834: contig of 8/6 bp in length
* 1835 1934: gap of 100 bp

1935 2832: contig of 898 bp in length
2833 2932: gap of 100 bp in length
2933 3785: contig of 857 bp in length
3790 3889: gap of 100 bp in length
3890 4780: contig of 891 bp in length
4781 4880: gap of 100 bp in length
4881 5740: contig of 860 bp in length
5741 5840: gap of 100 bp in length
5841 6742: contig of 902 bp in length
6743 6842: gap of 100 bp in length
6843 7684: contig of 842 bp in length
7685 7784: gap of 100 bp in length
7785 8609: contig of 825 bp in length
8610 8709: gap of 100 bp in length
8710 9573: contig of 864 bp in length
9574 9673: gap of 100 bp in length
9674 10537: contig of 864 bp in length
10538 10637: gap of 100 bp in length
10638 11521: contig of 884 bp in length
11522 11621: gap of 100 bp in length
11622 12466: contig of 845 bp in length
12467 12566: gap of 100 bp in length
12567 13464: contig of 898 bp in length
13465 13564: gap of 100 bp in length
13565 14419: contig of 855 bp in length
14420 14519: gap of 100 bp in length
14520 15368: contig of 849 bp in length
15369 15468: gap of 100 bp in length
15469 16325: contig of 857 bp in length
16326 16425: gap of 100 bp in length
16426 17261: contig of 836 bp in length
17262 17361: gap of 100 bp in length
17362 18185: contig of 824 bp in length
18186 18285: gap of 100 bp in length
18286 19098: contig of 813 bp in length
19099 19198: gap of 100 bp in length
19199 20057: contig of 859 bp in length
20058 20157: gap of 100 bp in length
20158 21021: contig of 864 bp in length
21022 21121: gap of 100 bp in length
21122 21995: contig of 874 bp in length
21996 22095: gap of 100 bp in length
22096 22965: contig of 870 bp in length
22966 23065: gap of 100 bp in length
23066 23957: contig of 892 bp in length
23958 24057: gap of 100 bp in length
24058 24926: contig of 869 bp in length
24927 25026: gap of 100 bp in length
25027 25881: contig of 855 bp in length
25882 25981: gap of 100 bp in length
25982 26827: contig of 846 bp in length
26828 26927: gap of 100 bp in length
26928 27791: contig of 864 bp in length
27792 27891: gap of 100 bp in length
27892 28742: contig of 851 bp in length
28743 28842: gap of 100 bp in length
28843 29689: contig of 847 bp in length
29690 29789: gap of 100 bp in length
29790 30642: contig of 853 bp in length
30643 30742: gap of 100 bp in length
30743 31640: contig of 898 bp in length
31641 31740: gap of 100 bp in length
31741 32602: contig of 862 bp in length
32603 32702: gap of 100 bp in length
32703 33554: contig of 852 bp in length
33555 33654: gap of 100 bp in length
33655 34521: contig of 867 bp in length
34522 34621: gap of 100 bp in length
34622 35483: contig of 862 bp in length
35484 35583: gap of 100 bp in length
35584 36455: contig of 872 bp in length
36456 36555: gap of 100 bp in length
36556 37414: contig of 859 bp in length

37415 37514: gap of 100 bp in length
37515 38355: contig of 841 bp in length
38356 38455: gap of 100 bp in length
38456 39298: contig of 843 bp in length
39299 39398: gap of 100 bp in length
39399 40234: contig of 836 bp in length
40235 40334: gap of 100 bp in length
40335 41179: contig of 845 bp in length
41180 41279: gap of 100 bp in length
41280 42148: contig of 869 bp in length
42149 42248: gap of 100 bp in length
42249 43083: contig of 835 bp in length
43084 43183: gap of 100 bp in length
43184 44082: contig of 899 bp in length
44083 44182: gap of 100 bp in length
44183 45028: contig of 846 bp in length
45029 45128: gap of 100 bp in length
45129 45993: contig of 865 bp in length
45994 46093: gap of 100 bp in length
46094 46970: contig of 877 bp in length
46971 47070: gap of 100 bp in length
47071 47976: contig of 906 bp in length
47977 48076: gap of 100 bp in length
48077 48933: contig of 857 bp in length
48934 49033: gap of 100 bp in length
49034 49927: contig of 894 bp in length
49928 50027: gap of 100 bp in length
50028 50871: contig of 844 bp in length
50872 50971: gap of 100 bp in length
50972 51816: contig of 845 bp in length
51817 51916: gap of 100 bp in length
51917 52755: contig of 839 bp in length
52756 52855: gap of 100 bp in length
52856 53707: contig of 852 bp in length
53708 53807: gap of 100 bp in length
53808 54687: contig of 880 bp in length
54688 54787: gap of 100 bp in length
54788 55651: contig of 864 bp in length
55652 55751: gap of 100 bp in length
55752 56608: contig of 857 bp in length
56609 56708: gap of 100 bp in length
56709 57576: contig of 868 bp in length
57577 57676: gap of 100 bp in length
57677 58531: contig of 855 bp in length
58532 58631: gap of 100 bp in length
58632 59474: contig of 843 bp in length
59475 59574: gap of 100 bp in length
59575 60408: contig of 834 bp in length
60409 60508: gap of 100 bp in length
60509 61396: contig of 888 bp in length
61397 61496: gap of 100 bp in length
61497 62325: contig of 829 bp in length
62326 62425: gap of 100 bp in length
62426 63255: contig of 830 bp in length
63256 63355: gap of 100 bp in length
63356 64201: contig of 846 bp in length
64202 64301: gap of 100 bp in length
64302 65155: contig of 854 bp in length
65156 65255: gap of 100 bp in length
65256 66110: contig of 855 bp in length
66111 66210: gap of 100 bp in length
66211 67072: contig of 862 bp in length
67073 67172: gap of 100 bp in length
67173 68036: contig of 864 bp in length
68037 68136: gap of 100 bp in length
68137 69021: contig of 885 bp in length
69022 69121: gap of 100 bp in length
69122 69982: contig of 861 bp in length

Query Match 3.8%; Score 57.2; DB 2; Length 162548;
Best Local Similarity 52.5%; Pred. No. 0.00025;
Matches 232; Conservative 0; Mismatches 186; Indels 22; Gaps 4;

OY	754	ATTCAATTCATATGAGCCATTCTTCTTTAAAGGCTCTCCAGAGAAATATATTGGAAC	813
Db	148621	AATCTAGCTCACCCTGTCACCACTCCAAAAGCATTCCTGGAGACTAAAATTGAAAC	148582
OY	814	ATTTCGTAATCTATCAAGTATCATCAATGTCTGCTCTGCCACCACTAG-ACCATC	872
Db	148561	ANCT---TATTGTGTGTGCAAATGATATATGCTCGATCTCCATTAATATATGAAGTC	148505
OY	873	CCCATGAATACAGACACTGCGCTTCTTAGTGTTCCTGTATGTCTGTCTGTGGTACATG	932
Db	148504	CACGAAAGTAGATATATCTATTTGTTTTTCTTCCCGTATCCCTTGCCTAGAACTGTG	148445
OY	933	TGATTAATGCTCTAATTAACACTCTGTGGAGCAATTGAATCATCAGATAGCACCCCTC	992
Db	148444	CTTGACACATAGTAGATAGCTCCAAATTAATGCTCGTAGATCAATCAAAATTCATTTTAG	148385
OY	993	TCTGAGAGGCATATCTCATGTGTTATCTCCCAAGACATGAGGTAAAGACATTAT---CCC	1048
Db	148384	TGAGAAATACATTAATTTCAATAGCATCATCCAAAAGCAGAGGTGTGACAGTGTCTCCC	148325
OY	1049	AGGTCATCTGCTGTTCGATATGATGATGTTCT-----AACCAAACTTAA	1094
Db	148324	AGATCCAAACTGGAATCTGTAATTCGTTCCCTATTTTTCATCTGCACAAATTAACTTAAAG	148285
OY	1095	TAGATTTAAAAACAGCAGGATTTATCTCACATGTTTTGAGACGCCAGAAATCTGCACACA	1154
Db	148264	TGGCTTAAAAACAACACAAACTTACTCTCTTAGAGTTCGAGAGCTAGAAATTAAGAAATCA	148205
OY	1155	GTTTCATCTTTAGACTGATG	1176
Db	148204	ATTCCACTGGACTTAAATCTACG	148183

[illegible]

COMMENT On Sep 29, 2001 this sequence version replaced gi:1356
----- Genome Center
Center: Multimegabase Sequencing Center
Center code: UMSC
Web site: http://chroma.mbt.washington.edu/msg_www
Contact: leeerwensystemsbiology.org

```
----- Summary Statistics -----
Sequencing vector: puc18; L08752
Chemistry: Dye-terminator Big Dye; 90% of reads
Chemistry: Dye-primer Big Dye; 10% of reads
Assembly program: Phrap; version 0.990399
```

Note: Data from overlapping BACS AC009996 (Drafting center: UWMSC) and AC018901 (Drafting center: UWMSC) were added for finishing.

FEATURES
Source

Location/Qualifiers
1. .128484

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
```

```
/map="15q15"
/clone="CTD-2165P21"
/clone_1fb="Cal Tech hum
```

RP11-292P13 AC018901. Data from overlapping BACs w

added and the consensus sequence determined. CTD-2165P21 to the extent possible." 1-87241

```
/note="overlap with RP11-151N17, AC009996"  
47494
```

```
/note="low quality data"  
51343. .51390  
/note="low quality data"
```

```
51502. .51594
/note="low quality data"
87236 128484
```

```
0/250.0:1.240404
/note="overlap with RP11-292P13, AC018901"
105444.105452
```

BASE COUNT	38664	a	27261	c	26033	g	36526	t
ORIGIN								

Query Match	3.5%	Score 53;	DB 9;	Length 128484;
Best Local Similarity	48.2%	Prod No	0.0032.	

Best Local	40.25	Freq: 0.00002	
Matches 149; Conservative	0	Mismatches 160	Indels 0; Gaps 0

0y	TCCTCATGACATCTCGAAATATTTCCAAATTTCTCTATTCGTGCAAAAGCTGAGGAAGTCTA	582
523	TCCTCATGACATCTCGAAATATTTCCAAATTTCTCTATTCGTGCAAAAGCTGAGGAAGTCTA	582
Db	TCCTTAATGTTGCCGGAATTTTACAGTGATTTTTCATGAGCCCATTTCTTTACCAACTGTA	66744
66803	TCCTTAATGTTGCCGGAATTTTACAGTGATTTTTCATGAGCCCATTTCTTTACCAACTGTA	66744

583 GTCCTATGAGATCTCCCTTTCCTCAGGTCCTCCGTGCTCCACACACCATTCACACTTCT 642

Db 66743 TTTTAAACCTCTCTGTTCCTCAAACTCCATTAATCTTTATTCCTTATTCCTCAACTTTTCT 66684

66683 TGGATGACCTTGGCCCTCATGAAGGCCCATCCTCGTGTTGTGCTTAGATCCATTCT 66624
oy 643 TGATATCTCTTGAACATAACAATTCCTCCATGGGTTGTCTCCCTTACCACAAATCA 702

703 TGCCTCAGGATCTACTGCCCCATCTTCACTCATCTCTGCTTGGTCATTCAATC 762

Db 66623 TCCATCCTTCTTAGGGGCGCTTCTCCATTTCCCTCTTTTATATAGTGTGCGCTTTA 66564

QY 763 TCATGTGCGCATTTCTAAAGCGCTCCACAGATATATATTGAACGATTTTGGTA 822

66563 TTTTCAGTTTCAAGTTTCTCCCCAGTCTTGTCCTTCCGGAAAATATATCATTAACTAAGTA 66504

[illegible]

Db 66503 TTTTAGAA 66495

```
RESULT 8
-----
```

AC009996/C	162495 bp	DNA	linear	PRI 28-APR-2001
LOCUS				
DEFINITION	Homo sapiens chromosome 15 clone RP11-151N17 map 15q15, complete			

sequence.	
ACCESSION	
AC009996	
VERSION	
AC009996.7	
GI	13876478

INDEXED
KEYWORDS
HTG.

ORGANISM	Homo sapiens.
SOURCE	Homo sapiens.
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 162495) Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S., Dots,M., Dichtoff,R., Harrison,G., James,R., Lasky,S., Burke,J., Dots,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R., Pate,D. and Hood,L.
TITLE	Sequencing of human chromosome 15 D15S146-D15S117 region
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 162495) Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B., Bloom,S., Dots,M., Dichtoff,R., Harrison,G., James,R., Lasky,S., Madan,A., Raccliff,A., Shaffer,T. and Hood,L.
AUTHORS	Direct Submission Submitted (10-SEP-1999) Multimegabase Sequencing Center, University of Washington, PO BOX 357730, Seattle, WA 98195, USA 3 (bases 1 to 162495) Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S., Burke,J., Dots,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R., Pate,D. and Hood,L.
TITLE	Direct Submission
JOURNAL	Submitted (22-APR-2001) Multimegabase Sequencing Center, Institute for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA 98105, USA
COMMENT	On Apr 28, 2001 this sequence version replaced gi:13621216. ----- Genome Center Center: Multimegabase Sequencing Center Center code: UWMSC Web site: http://chroma.mbt.washington.edu/msg_www Contact: leetowen@systemsbiology.org ----- Summary Statistics Sequencing vector: pUC18; 108752 Chemistry: Dye-terminator Big Dye; 90% of reads Chemistry: Dye-Primer Big Dye; 10% of reads Assembly program: Phrap; version 0.990399
FEATURES	Data from overlapping BACs AC025430 [Drafting center: UWMSC] and AC025270 [Drafting center: UWMSC] were added for finishing.
SOURCE	Location/Qualifiers 1. 162495 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="15" /map="15q15" /clone="RP11-151N17" /clone_lib="RPC1 human BAC library 11" /note="Data from overlapping BACs CTD-2329K10, CTD-2154L5 and CTD-2165P21 were added and the consensus sequence was determined from RP11-151N17 to the extent possible." 1. 4804 /note="Overlap with CTD-2329K10 AC025430" 29915..30270 misc_feature /note="sequence data generated from subcloned PCR product" 31515..31587 /note="low quality data" 35369..35617 /note="low quality data" 75258..162495 /note="Overlap with CTD-2165P21 AC025270" 90757..90786 /note="low quality data" 126567..126648 /note="low quality data" 126759..126851 /note="low quality data" BASE COUNT 48294 a 33902 c 34028 g 46271 t ORIGIN
Query Match	3.5%; Score 53; DB 9; Length 162495;
Best Local Similarity	48.2%; Pred. No. 0.0032;
Matches 149; Conservative	0; Mismatches 160; Indels 0; Gaps 0;

QY	523	TCCTATGACATCTGCAGAAATTTTCCAAATTTCTCATTTGCTGCAAGGTCTGGAGAGCTGA	582
Db	142038	TCCTATGTTGCTGCTGGAAATTTTACAGTGAATTTTTCATGAGCCCATTTCTTTACAACTGTA	1419999
QY	583	GTCCTATGATGATCCTCTTTTCTCCAGGTGTCCTGAGTCTCCACACACCAATTCACACTCT	642
Db	141998	TTTTTAACACTTCCTCTGTTTCTCCTCAAAACCCCAATTATCTTTATTTCCCAACTTTCT	1419393
QY	643	TGAATATTTCTTGGACATAACAATAATTCCTCCATGGGTTTGTTCCTCTACCCAAATTCGA	702
Db	141938	TGGAATGACCTTGCTGCTCTCTAATGAAGGCCCAATCCCTCTGTGTGCTTTGATATCCCAATTCCT	1418797
QY	703	TGCCCTTCAGAGTAATCTACTCTGCCCCATCTTCACTCATCTCTGCTTGTGTGCTATTAATAATC	762
Db	141878	TCCATTCCTCTCTAAGGGCCCTTCTTCATTTCTCTCTTTTATATAATGTTGTGCCCTTGA	1418195
QY	763	TCAAAATGTAAGCAATTTCTTAAGAGGCTCTCCAGAGAAATATATTGGAAGCAATTTTGCTA	822
Db	141818	TCCACAGTTTACAGTTTCTCCCAAGTCAAGTTGTGTTCCGGAAATATTTACATACAGTAAGATA	1417599
QY	823	TTCTATTCAA	831
Db	141758	TTTTAAGAA	141750

RESULT 9	AL359543/c	AL359543	98151 bp	DNA	linear	PRI 30-JAN-2002
LOCUS						
DEFINITION		Human DNA sequence from clone Rp11-9801 on chromosome 1, complete sequence.				
ACCESSION		AL359543				
VERSION		AL359543.13	GI:18476597			
KEYWORDS		HTG.				
SOURCE		human.				
ORGANISM		Homo sapiens				
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE		1 (bases 1 to 98151)				
AUTHORS		Brown, A.				
TITLE		Direct Submission				
JOURNAL		Submitted (30-JAN-2002)				
		Wellcome Trust Sanger Institute, Hinxton,				

COMMENT

hunqurey@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Feb 1, 2002 this sequence version replaced q1:18250465.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw:
SWISSPROT; Tr: TrEMBL; Wp: WormPEP. Information on the WormPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
Chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr1>
Rp11-9601 is from the library RPCR-11.1 constructed by the group of
Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
VECTOR: pBAC3.6
IMPORTANT: This sequence is not the entire insert of clone
Rp11-9601. It may be shorter because we sequence overlapping
sections only once, except for a short overlap.

FEATURES
source

The true left end of clone RP11-9801 is at 1 in this sequence. The true left end of clone RP11-555124 is at 96152 in this sequence. The true right end of clone RP11-17H4 is at 37904 in this sequence.

1. 98151

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="1"

/clone="RP11-9801"

/clone_1lb="RPCT-11.1"

BASE COUNT 25770 a 20210 c 21225 g 30946 t

ORIGIN

Query Match 3.5%; Score 52.6; DB 9; Length 98151;

Best Local Similarity 77.1%; Pred. No. 0.0042;

Matches 64; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

OY 1081 TAACACAACTTAATGATTAAACACAGCATTTTCTGCACATGTTTGACAGCCCA 1140

DB 83800 TACCAACAACTTACAGGCTTAAACACATTAATTTTCTGCTCATGTTTACAGCCA 83741

OY 1141 GAAATCTGACACACATTTTCATG 1163

DB 83740 GAAGTCTGAATCATGTTTACTG 83718

RESULT 10

AC015524 178985 bp DNA linear PRI 01-MAY-2001

LOCUS Homo sapiens chromosome 18, clone RP11-202D1, complete sequence.

AC015524

AC015524.7 GI:13899436

KEYWORDS HTG.

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 178985)

2 (bases 1 to 178985)

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

TITLE
JOURNAL

COMMENT

Submitted (01-MAY-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On May 1, 2001 this sequence version replaced gi:12621436.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L4275

Center clone name: 202_D_1

FEATURES

source

Location/Qualifiers

1. 178985

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="18"

/map="18"

/clone="RP11-202D1"

/clone_1lb="RPCT-11 Human Male BAC"

1. 178985

/note="<30 qual SNGL region"

1. 127

/rpt_family="LM"

1082. 1253

/rpt_family="MSTB"

1254. 1551

/rpt_family="Alusx"

1552. 1742

/rpt_family="MSTB"

1772. 3186

/rpt_family="MSTB-int"

3187. 3368

/rpt_family="MSTB"

3369. 3395

/rpt_family="MSTB"

3396. 3574

/rpt_family="MSTB"

3555. 3757

/rpt_family="MSTB"

3757. 3918

/rpt_family="MSTB"

3918. 4079

/rpt_family="MSTB"

4079. 4240

/rpt_family="MSTB"

4240. 4401

/rpt_family="MSTB"

4401. 4562

/rpt_family="MSTB"

4562. 4723

/rpt_family="MSTB"

4723. 4884

/rpt_family="MSTB"

4884. 5045

/rpt_family="MSTB"

```

repeat_region      /rpl_family="Tiger2a"
                    13283..13344
repeat_region      /rpl_family="MER103"
                    13656..13999
repeat_region      /rpl_family="LIPa8"
                    14772..15216
repeat_region      /rpl_family="MT1C"
                    15224..15458
repeat_region      /rpl_family="MIR"
                    15913..15936
repeat_region      /rpl_family="AT-rich"
                    complement(16513..16856)
repeat_region      /rpl_family="MT1A1"
                    17189..17274
repeat_region      /rpl_family="(TATATG)n"
                    17503..17820
repeat_region      /rpl_family="AluDb"
                    complement(19008..19327)
repeat_region      /rpl_family="MER102b"
                    complement(19595..19782)
repeat_region      /rpl_family="MT1H1"
                    complement(19816..20030)
repeat_region      /rpl_family="MT1H"
                    complement(20227..20387)
repeat_region      /rpl_family="LIMEa"
                    20433..20589
repeat_region      /rpl_family="MER5a"
                    21446..21488
repeat_region      /rpl_family="(TG)n"
                    complement(22330..22471)
repeat_region      /rpl_family="MIR"
                    22657..22945
repeat_region      /rpl_family="HALb"
                    23077..23192
repeat_region      /rpl_family="LIPa5"
                    23363..23389
repeat_region      /rpl_family="(TA)n"
                    24013..24430
repeat_region      /rpl_family="MSTB2"
                    24802..24857
repeat_region      /rpl_family="A-rich"
                    25010..25596
repeat_region      /rpl_family="LIMC4a"
                    25836..26076
repeat_region      /rpl_family="LIMC4a"
                    26227..26486
repeat_region      /rpl_family="LIMC4a"
                    26497..26626
repeat_region      /rpl_family="LIMC4a"
                    26761..26819
repeat_region      /rpl_family="LIM1"
                    26879..27200
repeat_region      /rpl_family="LIMC4a"
                    27219..27441
repeat_region      /rpl_family="AluDb"
                    complement(27694..27765)
repeat_region      /rpl_family="MIR"
                    28059..28233
repeat_region      /rpl_family="MER5a"
                    28538..28790
repeat_region      /rpl_family="LIM2"
                    28980..29044
repeat_region      /rpl_family="CT-rich"
                    29141..29207
repeat_region      /rpl_family="L3"
                    29246..29505
repeat_region      /rpl_family="LIM9a"
                    29514..29649
repeat_region      /rpl_family="AluDb"
                    29658..29793
repeat_region      /rpl_family="LIM9a"
                    30147..30204
repeat_region      /rpl_family="(TA)n"

```

```

repeat_region      30294..30332
                    /rpl_family="(TATATG)n"
repeat_region      complement(30333..30650)
                    /rpl_family="LIPa8"
repeat_region      31149..31487
                    /rpl_family="Tiger2a"
repeat_region      complement(31488..31779)
                    /rpl_family="AluDb"
repeat_region      31780..31917

Query Match      3.4% Score 51.4; DB 9; Length 178985;
Best Local Similarity 49.1%; Pred. No. 0.0083;
Matches 136; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 1005 GATTCATGCTTATCCCAAGACATGAGTAAAGACATTAATCCAGCTGCCTGTTT 1064
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 72476 GATTCACACACGCACTCCAGCTGGAGATTAAGCAGACATCCCATCATACCTTCT 72535
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1065 CCGTATTGTTTCTTAACACAACTTAATAGATTAAACAGCAGGATTATTCAC 1124
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 72536 GGCCTCTGAATGAATTAACCATCTGTGATGATTCMAACAGATATTATTCCTC 72595
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1125 ATGTTTAGAGCCAGAAATGACACCATTCATGTTGATGATGATGACACACTG 1184
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 72596 ACAGTTCTAGAGTCGACAGATCCAGATGATGTTCACTGCTTAAGATCAAGTGTCAGC 72655
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1185 TAATCTGCTACTTAGAGCAGATGACGGGAGCATATGATTAAGCCCATTTTAAAGC 1244
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 72656 AGGCCATTTTCCCTCTATAGGCTCTCAGGAGAAATCTGATTTATTCCTACTCTCCGG 72715
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1245 TCGTGGGTGAGAACCTGCTCTGATTTTTCACAT 1281
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 72716 GTTGGTGGCTGCCAAGATGATGCTTTGGCCT 72752
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

RESULT 11
AC021757/c 201156 bp DNA linear HTG 24-MAY-2001
LOCUS Homo sapiens clone RP11-45M11, WORKING DRAFT SEQUENCE, 7 unordered
DEFINITION
ACCESSION AC021757
VERSION AC021757.5 GI:14192966
KEYWORDS HTG; HTGS-PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 201156)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerley,R., Beda,F.,
Boguslavsky,L., Bouckgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepe,Y., Colangelo,M., Collins,S., Collimore,A., Cooke,P.,
DeRellano,K., Dewar,K., Domino,M., Doyle,M., Fenebor,J.,
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,U.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheters,R., Meldrum,J., Menus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Tirell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
JOURNAL Direct Submission
TITLE Submitted (20-JAN-2000) Whitehead Institute/MIT Center for Genome
JOURNAL Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On May 24, 2001 this sequence version replaced gi:13259440.

```



```

repeat_region      16655..17070          /rpt_family="MLR7D"
repeat_region      complement(17454..17763)
repeat_region      /rpt_family="AluY"
repeat_region      complement(17815..17922)
repeat_region      /rpt_family="FLW_M_C"

Query Match              3.4%; Score 50.4; DB 9; Length 167746;
Best Local Similarity    75.0%; Pred. No. 0.015;
Matches   63; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

OY      1141 GAATCTGCACACCAGTTTCATGT 1164
Db 148168 GAAGTCTGACATCAGTTTCACGTCT 148145

RESULT 13
AC113567
LOCUS
DEFINITION
Canis familiaris clone Rpel-265M12, WORKING DRAFT SEQUENCE, 2
unordered pieces.
AC113567
VERSION
AC113567.1 GI:19071598
KEYWORDS
HTGS: HTGS_PHASE1: HTGS_DRAFT.
SOURCE
Canis familiaris.
ORGANISM
Eumariota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE
Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Deutch,N.L., Grantte,S., Guan,X., Gupta,V., Haghighi,P.,
Ho,S.-L., Idol,J.R., Karlins,E., Latic,P., Lee-Lin,S.-Q.,
Legaspi,R., Maduro,Q.L., Maduro,A.B., Mastello,C., Maskeri,B.,
Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Prasad,A.,
Stantilopos,S., Thomas,J.W., Thomas,P.J., Touchman,J.W.,
Tsurgoun,C., Vogt,J.L., Walker,M.A., Wecherby,K.D., Wiggins,L.,
Young,A., Zhang,L.H. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2. (bases 1 to 172884)
Green,E.D.
TITLE
Direct Submission
JOURNAL
Submitted (02-MAR-2002) NIH Intramural Sequencing Center, 8717
AUTHORS
Glovermont Circle, Gaithersburg, MD 20877, USA
JOURNAL
Genome Center
COMMENT
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_mouse@nigrl.nih.gov
Project Information
Center project name: c1o
Center clone name: 265M12
Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 172016 bases at least Q40
Consensus quality: 172086 bases at least Q30
Consensus quality: 172156 bases at least Q20
Insert size: 134000; agarose-fp
Insert size: 172784; sum-of-contigs
Quality coverage: 12.36x in Q20 bases; agarose-fp
Quality coverage: 9.59x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record as
* arbitrary. Gaps between the contigs are represented as

```

```

* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 77918: contig of 77918 bp in length
* 77919 78018: gap of unknown length
* 78019 172884: contig of 94866 bp in length.
FEATURES
    source
        1..172884
            /organism="Canis familiaris"
            /db_xref="taxon:9615"
            /clone="RP81-265M12"
            /clone_1bp="RP81"
            1..77918
                /note="assembly-fragment
                clone_end:SP6
                vector_side:left"
            78019..172884
                /note="assembly-fragment
                clone_end:T7
                vector_side:right"
BASE COUNT      56401 a 31413 c 30627 g 54342 t 101 others
ORIGIN
Query Match      3.4% Score 50.4; DB 2; Length 172884;
Best Local Similarity 69.0%; Pred. No. 0.015;
Matches 69; Conservative 0; Mismatches 31; Indels 0; Gaps 0.
Oy 1077 TTCTCAACAAACACTTATAGATTAAACAGACGATTTATCTCACATCTTTGAGAC 1136
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36420 TTCCACACCAACAACCTGGTAGCTTAAACACCAAAAATTTATCTTTTCAACATCTGGAG 36479
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 1137 GCCAGAACTCGACACGCTTTCATCTTGAAGCTTGATG 1176
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36480 GCCAGAACTCGAAGACGATTTCACTGCACGTAAGTAAG 36519
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
AC117938      193209 bp      DNA      linear      HTG 11-APR-2002
LOCUS      AC117938
DEFINITION      Canis familiaris clone Rp81-37219, WORKING DRAFT SEQUENCE, 2
unordered pieces.
AC117938
AC117938.1 GI:20136912
HTG: HTGS-PHASE1; HTGS-DRAFT.
SOURCE      Canis familiaris
ORGANISM      Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE      1 (bases 1 to 193209)
AUTHORS      Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Stenberg,S.M.,
Benjamin,L., Blackley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
Haghighi,P., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Latic,P.,
Lee-Lin,S.-O., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
Marquis,E.H., Mastello,C., Maskeri,B., Mastrian,S.D.,
McLoakey,J.C., McDowell,J., Paguligan,C., Pearson,R.,
Portnoy,M.E., Prasad,A., Schuele,M.G., Stantilpop,S., Thomas,D.W.,
Thomas,P.J., Touchman,J.W., Tsugeon,C., Vogt,J.L., Walker,M.A.,
Wehedy,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 193209)
Green,E.D.
Direct Submission
Submitted (11-APR-2002) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaitersburg, MD 20877, USA
Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoehngrl.nih.gov
----- Project Information -----

```

```

----- Summary Statistics -----
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 192769 bases at least Q40
Consensus quality: 192942 bases at least Q30
Consensus quality: 193052 bases at least Q20
Insert size: 154000; agarose-fp
Insert size: 193109; sum-of-contigs
Quality coverage: 11.42x in Q20 bases;
Quality coverage: 9.11x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*      1      8872: contig of 8872 bp in length
*      *      8873      8972: gap of unknown length
*      *      8973      193209: contig of 184237 bp in length.
FEATURES             source
source
    1..193209
        /organism="Canis familiaris"
        /db_xref="taxon:9615"
        /clone="RP81-37219"
        /clone_1fb="RP81"
    1..8872
        /note="assembly_fragment"
        clone_end:sp6
        vector_side:right
    8973..193209
        /note="assembly_fragment"
        clone_end:r7
        vector_side:right"
BASE COUNT          62568 a 34671 c 34070 g   61800 t   100 others
ORIGIN
Query Match              3.4%; Score 50.4; DB 2; Length 193209;
Best Local Similarity    69.0%; Pred. No. 0.015;
Matches 69; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
QY 1077 TTTCTAACAACAACATTAATAGATTAAACAGACGAGTATTATTCACACTGTGGAGAC 1136
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 184375 TTCCCAACACACAACTTGTAAGTTAAACACACAAAATTTATCTTTTTCACATTTCTGGAG 184434
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1137 GCCAGAAATCTGACACACAGTTTCAATGTTTAGACTTGANG 1176
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 184435 GCCAGAATCTGAAGCAGTTTCACTGCACGTCAAGTAAG 184474
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
LOCUS               AC003667                134019 bp     DNA           linear       PRI 29-MAR-2002
DEFINITION          Homo sapiens Xp22 PAC Research Park Cancer Institute PAC library)
SOURCE              AC003667
KEYWORDS            complete sequence.
ACCESSION            AC003667
VERSION             AC003667.2 GI:18921262
XREF                 HTG.
ORGANISM            Homo sapiens.
                     Homo sapiens.
                     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                     Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE            1 (bases 1 to 134019)
AUTHORS             Muny,D., Arenson,A.D., Brundage,E., Carvelli,K., Chen,E., Di,W.,
                     Ding,X., Dugan,S., Durbin,J., Forcum,J., Ganesh,R., Garcia,C.,
                     Goodman,M., Gorrell,J.H., Haywood,M., Jackson,L., Kampal,R.,
                     Karpathy,S., Leal,B., Liu,W., Logan,O., Lu,J., Ly,T., Martinez,C.,
                     Oswal,G., Perez,L., Rashid,N.D., Rowland,K., Savage,L.,

```

```

TITLE                               Scherer,S.S., Shen,H., Tims,K.M., Todd,J., Vo,Q., Worley,K.C.,
JOURNAL                             Xu,W., Chinault,C., Nelson,D. and Gibbs,R.A.
REFERENCE                           Direct Submission
AUTHORS                             2 (bases 1 to 134019)
TITLE                               Unpublished
JOURNAL                             Worley,K.C.
REFERENCE                           Direct Submission
AUTHORS                             Submitted (09-DEC-1997) Molecular and Human Genetics, Baylor
TITLE                               College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
JOURNAL                             3 (bases 1 to 134019)
AUTHORS                             Worley,K.C.
TITLE                               Direct Submission
REFERENCE                           Submitted (20-FEB-1998) Molecular and Human Genetics, Baylor
JOURNAL                             College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REMARK                             Completed sequence
REFERENCE                           4 (bases 1 to 134019)
AUTHORS                             Worley,K.C.
TITLE                               Direct Submission
JOURNAL                             Submitted (28-MAR-1998) Human Genome Sequencing Center, Department
REMARK                             of Molecular and Human Genetics, Baylor College of Medicine, One
AUTHORS                             Baylor Plaza, Houston, TX 77030, USA
TITLE                               5 (bases 1 to 134019)
JOURNAL                             Worley,K.C.
REFERENCE                           Direct Submission
AUTHORS                             Submitted (01-APR-1998) Human Genome Sequencing Center, Department
TITLE                               of Molecular and Human Genetics, Baylor College of Medicine, One
JOURNAL                             Baylor Plaza, Houston, TX 77030, USA
AUTHORS                             6 (bases 1 to 134019)
TITLE                               Worley,K.C.
REFERENCE                           Direct Submission
AUTHORS                             Submitted (26-FEB-2002) Molecular and Human Genetics, Baylor
TITLE                               College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
JOURNAL                             Completed sequence
REMARK                             7 (bases 1 to 134019)
AUTHORS                             Worley,K.C.
TITLE                               Direct Submission
JOURNAL                             Submitted (29-MAR-2002) Molecular and Human Genetics, Baylor
REMARK                             College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
AUTHORS                             Completed sequence
TITLE                               On Feb 26, 2002 this sequence version replaced g1:2995482.
JOURNAL                             Sequencing is completed to a minimum standard of double strand
REMARK                             coverage with a minimum of 2 clones and 2 reads with no ambiguities
AUTHORS                             or 2 chemistries with a minimum of 2 clones and 3 reads with no
TITLE                               ambiguities. If the sequence quality does not meet this standard,
JOURNAL                             it will be indicated in the annotation.
REMARK                             The repeat regions shown were identified using RepeatMasker by
COMMENT                             Adrian Smlt.
                                     Sequence similarities were identified using Powerblast by Jinghui
                                     Zhang.
Exon/Intron boundaries of identified genes were chosen if there
were canonical splice junctions that maintained sequence continuity
across the splice junctions.
Location/Qualifiers
1. .134019
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="X"
   /map="Xp22"
   /clone="RPC11-17L20"
   /clone_1lb="Research Park Cancer Institute PAC library"
   complement(1..2023)
   /note="overlaps bases 1,2023 of clone AC097348"
   /function="clone overlap"
complement(1..2004)
   /note="overlaps bases 1..2004 of clone AC112916"
   /function="clone overlap"
8..89
   /rpt_family="MER1B"
   /rpt_region
86..181
   /rpt_family="MER21B"

```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 16, 2003, 23:50:11 ; Search time 280.686 Seconds

(without alignments)
12034.813 Million cell updates/sec

Title: US-09-605-042a-1_COPY_1_1500
Perfect score: 1500

Sequence: 1 gggggggggccctcgagggtt.....agatcccccattctctctcg 1500

Scoring table: IDENTITY_NTC
Gapex 10.0 , Gapex 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq.101002.*
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1500	100.0	9345	21	AA293966
2	44.6	3.0	472	22	ABA57262
3	44.6	3.0	472	22	AAK05289
4	44.6	3.0	472	22	AAK30881
5	44.6	3.0	472	22	AAK136794
6	44.6	3.0	472	22	ABSO5629
7	44.6	3.0	579	22	ABA63337
8	43.6	2.9	1503900	22	AAK95240
9	43.6	2.9	1503900	22	AAK96733

C	10	43.2	2.9	585	22	ABA60114	Human foetal liver
C	11	43.2	2.9	585	22	ABA28474	Probe #6940 for ge
C	12	43.2	2.9	585	22	AAK08390	Human brain expres
C	13	43.2	2.9	585	22	AAK34270	Human bone marrow
C	14	43.2	2.9	585	22	AAI39993	Probe #8679 used t
C	15	43.2	2.9	585	22	ABSO8999	Human genome-deriv
C	16	43	2.9	1750	21	AAI18143	Lung cancer associ
C	17	43	2.9	2747	22	ABV24568	Human prostate exp
C	18	43	2.9	28180	22	AAK68939	Human immune/haema
C	19	42.2	2.8	401	22	AAK95381	Human neurogulin g
C	20	42.2	2.8	401	22	AAK96874	Human ovarian and
C	21	41.4	2.8	31813	22	ABA08137	Human reproductive
C	22	41.4	2.8	31813	22	AAI06997	Gene #1543 used to
C	23	41.4	2.8	55795	22	ABN95045	Kidney cancer rela
C	24	41.4	2.8	55795	22	ABN95045	Kidney cancer rela
C	25	41.4	2.8	55795	22	ABL68242	Kidney cancer rela
C	26	41.4	2.8	55795	22	ABL68484	Kidney cancer rela
C	27	40.6	2.7	478	21	AAI15173	Human genome-deriv
C	28	40.6	2.7	543	22	ABA62656	Human secreted pro
C	29	40.6	2.7	543	22	ABA29965	Human foetal liver
C	30	40.6	2.7	543	22	AAK11028	Probe #8431 for ge
C	31	40.6	2.7	543	22	AAK36862	Human brain expres
C	32	40.6	2.7	543	22	AAI17715	Human bone marrow
C	33	40.6	2.7	543	22	AAI42652	Probe #11338 used
C	34	40.6	2.7	543	22	ABSI0874	Human genome-deriv
C	35	40.6	2.7	1479	20	AAZ17259	Human gene express
C	36	40.6	2.7	24768	22	AAK91299	Human digestive sy
C	37	40.6	2.7	24768	22	AAK91299	Human liver associ
C	38	40.6	2.7	24768	22	ABN90481	Human liver associ
C	39	40.4	2.7	50000	21	AAK63672	Polymorphic repeat
C	40	40.4	2.7	50000	21	AAK63672	Nucleotide sequenc
C	41	40.2	2.7	6603	22	AAK64182	Human immune/haema
C	42	39.8	2.7	52354	24	AAK35032	Human transporter
C	43	39.8	2.7	378	22	AAI91377	Human polynucleoti
C	44	39.8	2.7	2604	21	AACT7400	Human secreted pro
C	45	39.8	2.7	6741	21	AAAI0595	Gene encoding a su

ALIGNMENTS

RESULT 1
AA293966 standard; DNA; 9345 BP.

XX AA293966;

DT 29-AUG-2000 (first entry)

XX Mouse uromodulin promoter.

DE Uromodulin; promoter; kidney; urine; heterologous gene; treatment;

KW therapy; gene expression; pharmaceutical; mouse; ds.

RW Mus musculus.

XX Mus musculus.

OS Mus musculus.

FT Key Location/Qualifiers

FT misc-feature 4..5

FT exon /tag= a

FT intron /label= Polycloning site of pbs

FT /cons_splice= (5'site:NO, 3'site:YES)

FT /tag= b

FT /tag= c

FT /label= Intron 1

FT /tag= d

FT /label= Exon 2

FT /con_splice= (5'site:NO, 3'site:YES)

FT /tag= e

FT intron

FT Human

FT misc-feature /label= Intron 2
8273..8274
FT exon /tag= f
FT exon /note= "APA/RPN Junction"
FT exon /tag= g
FT exon /label= Exon 3
XX WO200029608-A1.
XX PD 25-MAY-2000.
XX 12-NOV-1999; 99MO-US26870.
XX 13-NOV-1998; 98US-0108195.
XX 09-JUL-1999; 99US-0142925.
XX (UYNV) UNIV NEW YORK STATE.
XX Wu X, Sun T;
PI WPI; 2000-387816/33.
XX New kidney-specific promoter useful for production of transgenic
PT animals as urinary bioreactors, is operably linked to a heterologous
PT gene
XX Claim 3; Fig 7a-7h; 55pp; English.
XX New methods to produce heterologous recombinant proteins in urine
CC require the use of a DNA molecule which is a kidney-specific
CC promoter, such as the uromodulin promoter, operably linked to a
CC heterologous gene encoding a biologically active protein. The
CC uromodulin promoter expresses the heterologous gene in vivo in the
CC kidneys to produce a recombinant biologically active protein in the
CC urine. The recombinant proteins produced may be useful for treating
CC human diseases. The major advantages of using this urine-based system
CC over milk-based systems are the ability to harvest the product soon
CC after birth and throughout the life of the animal irrespective of sex
CC or reproductive status, and the ease of product purification from
CC urine. In addition, livestock urine is a proven, currently utilized
CC source of pharmaceuticals.
XX

SO Sequence 9345 BP; 2622 A; 2072 C; 2084 G; 2565 T; 2 other:
Query Match 100.0%; Score 1500; DB 21; Length 9345;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGGGGGCCCTCGGGAGTTGGCTAGTCTGCAATGAGCTGTGATGACAGTTGGCG 60
DB 1 GGGGGGGCCCTCGGGAGTTGGCTAGTCTGCAATGAGCTGTGATGACAGTTGGCG 60
QY 61 CATATGAGATCCAGTACAGAGTCTCTAGATGTCATACCAATAGTACGCCATC 120
DB 61 CATATGAGATCCAGTACAGAGTCTCTAGATGTCATACCAATAGTACGCCATC 120
QY 121 ATTATGCAATCAGCGGAGTCTCTGCTGCTTGTCTTACTAGTGAACCTGA 180
DB 121 ATTATGCAATCAGCGGAGTCTCTGCTGCTTGTCTTACTAGTGAACCTGA 180
QY 181 TAACCTATATGATTTTACCAATTTCCCTCCATGGCAGTCACTCTCTTCTATG 240
DB 181 TAACCTATATGATTTTACCAATTTCCCTCCATGGCAGTCACTCTCTTCTATG 240
QY 241 ACCCTACTATGTCCTAGTGAAGTCTGCTCTTGTGATGAGAGCCATCTGTTCT 300
DB 241 ACCCTACTATGTCCTAGTGAAGTCTGCTCTTGTGATGAGAGCCATCTGTTCT 300
QY 301 TCTATGAGACTCTGCTACTCTTCTCCACGTGATCCACCAATCTGCTACATTGCA 360
DB 301 TCTATGAGACTCTGCTACTCTTCTCCACGTGATCCACCAATCTGCTACATTGCA 360

QY 361 CACTCAGAGTTCTTGGAGAGCAGAAAGACTCAGACTGATCTGCCCAATGCTCCCTA 420
DB 361 CACTCAGAGTTCTTGGAGAGCAGAAAGACTCAGACTGATCTGCCCAATGCTCCCTA 420
QY 421 CACTTCTCCATTAATCAGATATCTAAGCTATAGAGATTAATTCATGCACTAAGCT 480
DB 421 CACTTCTCCATTAATCAGATATCTAAGCTATAGAGATTAATTCATGCACTAAGCT 480
QY 481 TTCAGTACTATGATCTACTGTCCTACCCCTGTAACCTGATCTTCATGACATCTCGAA 540
DB 481 TTCAGTACTATGATCTACTGTCCTACCCCTGTAACCTGATCTTCATGACATCTCGAA 540
QY 541 TATTTCAATTTCTATGCTGCAAGCTGAGAGCTGATGATGATGATCTGCTT 600
DB 541 TATTTCAATTTCTATGCTGCAAGCTGAGAGCTGATGATGATGATGATCTGCTT 600
QY 601 TCTCTCAGAGTCTCTGCTGCTCAGACACACATTCACACTCTTGAATATCTTGAACAT 660
DB 601 TCTCTCAGAGTCTCTGCTGCTCAGACACACATTCACACTCTTGAATATCTTGAACAT 660
QY 661 AACAAATTCCTCCATGAGGTTGTTCCCTCTACCCAAATTCATGCTCAGGATCTTAC 720
DB 661 AACAAATTCCTCCATGAGGTTGTTCCCTCTACCCAAATTCATGCTCAGGATCTTAC 720
QY 721 TCTGCCCATCTTACATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
DB 721 TCTGCCCATCTTACATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
QY 781 AAAAGCTCTCCAGACAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 840
DB 781 AAAAGCTCTCCAGACAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 840
QY 841 CAATGCTGCTCTCCAGACACATTCATCCCATGATGATGATGATGATGATGATGATGAT 900
DB 841 CAATGCTGCTCTCCAGACACATTCATCCCATGATGATGATGATGATGATGATGATGAT 900
QY 901 TGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 960
DB 901 TGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 960
QY 961 GAGCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1020
DB 961 GAGCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1020
QY 1021 CCAAGCATGAGTGAAGATTAATCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGT 1080
DB 1021 CCAAGCATGAGTGAAGATTAATCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGT 1080
QY 1081 TAACCAAACTTAATGATTAATTAACAGCAGGATTAATTCATGATGATGATGATGATG 1140
DB 1081 TAACCAAACTTAATGATTAATTAACAGCAGGATTAATTCATGATGATGATGATGATG 1140
QY 1141 GAAATCTGACACACATTTCAATGTTTGAATCTGATGATGATGATGATGATGATGATG 1200
DB 1141 GAAATCTGACACACATTTCAATGTTTGAATCTGATGATGATGATGATGATGATGATG 1200
QY 1201 GAGGAGATGAGAGGAGTATGATTAAGCCCATTTTAACTGCTGCTGCTGCTGCTGCT 1260
DB 1201 GAGGAGATGAGAGGAGTATGATTAAGCCCATTTTAACTGCTGCTGCTGCTGCTGCT 1260
QY 1261 GCTGATGATTTTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1320
DB 1261 GCTGATGATTTTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1320
QY 1321 GGAAGAAACCTTTGCTTGCACCTTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1380
DB 1321 GGAAGAAACCTTTGCTTGCACCTTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1380
QY 1381 GGTCTGGAATCAGTGAAGCTTATGCTCAGTCCATTCATTCCTCTGATTTATCTCT 1440
DB 1381 GGTCTGGAATCAGTGAAGCTTATGCTCAGTCCATTCATTCCTCTGATTTATCTCT 1440
QY 1441 AAGCAGCGGTGCTTGTATCCAACTTTAAGAGCCCATAGATCCCATTTCTCTCTG 1500

DB 1441 AAGCAGCGGTGCTGTATCCACACTTTAGAGCCCATATGATCCCATTTCTCTCG 1500

RESULT 2

ABAS7262 standard; DNA; 472 BP.

ABAS7262;

01-FEB-2002 (first entry)

Human foetal liver single exon nucleic acid probe #5567.

Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

Homo sapiens.

WO200157277-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US00669.

04-FEB-2000; 2000US-0180312.

26-MAY-2000; 2000US-0207456.

30-JUN-2000; 2000US-0608408.

03-AUG-2000; 2000US-0632366.

21-SEP-2000; 2000US-0234687.

27-SEP-2000; 2000US-0236359.

04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI: 2001-483447/52.

Human genome-derived single exon nucleic acid probes useful for

analyzing gene expression in human fetal liver -

Claim 1; SEQ ID NO 5567; 639pp + sequence listing; English.

The invention relates to a single exon nucleic acid probe for

measuring human gene expression in a sample derived from human foetal

liver. The single exon nucleic acid probes may be used for predicting,

measuring and displaying gene expression in samples derived from human

fetal liver. The present sequence is a single exon nucleic acid

probe of the invention.

Note: The sequence data for this patent did not form part of the

printed specification, but was obtained in electronic format directly

from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 472 BP; 101 A; 102 C; 101 G; 168 T; 0 other;

Query Match 3.0%; Score 44.6; DB 22; Length 472;

Best Local Similarity 65.7%; Pred. No. 0.0035;

Matches 65; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

DB 1051 GTCCATGCTGTTCCGATGATTTCTTAACCAAACTTAATGATTAACAGCAC 1110

DB 164 GTACTGATTTTCAAGGCTTCTGTAGATTCACCAAACTTGCTTAACCAACAG 223

OY 1111 GGATTATTCACATGTTTGAAGAGCCAGAAATCTGA 1149

DB 224 AAATTATTCCTCAGAGTTCTAGAGGCCAGAAATCTGA 262

RESULT 3

AAK05289

ID AAK05289 standard; DNA; 472 BP.

XX AAK05289;

XX 05-NOV-2001 (first entry)

Human brain expressed single exon probe SEQ ID NO: 5280.

Human; brain expressed exon; gene expression analysis; probe;

microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;

epilepsy; cancer; ss.

Homo sapiens.

WO200157275-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US00667.

04-FEB-2000; 2000US-0180312.

26-MAY-2000; 2000US-0207456.

30-JUN-2000; 2000US-0608408.

03-AUG-2000; 2000US-0632366.

21-SEP-2000; 2000US-0234687.

27-SEP-2000; 2000US-0236359.

04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI: 2001-483446/52.

Single exon nucleic acid probes for analyzing gene expression in human

brains -

Example 4; SEQ ID NO: 5280; 650pp + Sequence Listing; English.

The present invention provides a number of single exon nucleic acid

probes which are derived from genomic sequences expressed in the human

brain. They can be used to measure gene expression in brain cell samples,

which may enable the diagnosis and improved treatment of nervous system

diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,

epilepsy and cancers. The present sequence is one of the probes of the

invention.

Sequence 472 BP; 101 A; 102 C; 101 G; 168 T; 0 other;

Query Match 3.0%; Score 44.6; DB 22; Length 472;

Best Local Similarity 65.7%; Pred. No. 0.0035;

Matches 65; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

OY 1051 GTCCATGCTGTTCCGATGATTTCTTAACCAAACTTAATGATTAACAGCAC 1110

DB 164 GTACTGATTTTCAAGGCTTCTGTAGATTCACCAAACTTGCTTAACCAACAG 223

OY 1111 GGATTATTCACATGTTTGAAGAGCCAGAAATCTGA 1149

DB 224 AAATTATTCCTCAGAGTTCTAGAGGCCAGAAATCTGA 262

RESULT 4

AAK30881

ID AAK30881 standard; DNA; 472 BP.

XX AAK30881;

XX 06-NOV-2001 (first entry)

DE Human bone marrow expressed single exon probe SEQ ID NO: 5438.

XX Human; bone marrow expressed exon; gene expression analysis; probe;

KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.

XX Homo sapiens.

```

XX  MO200157276-A2.
PN
XX
XX  09-AUG-2001.
XX
XX  30-JAN-2001; 2001WO-US00668.
XX
XX  04-FEB-2000; 2000US-0180312.
PR  26-MAY-2000; 2000US-0207456.
PR  30-JUN-2000; 2000US-0608408.
PR  03-AUG-2000; 2000US-0632366.
PR  21-SEP-2000; 2000US-0234687.
PR  27-SEP-2000; 2000US-0236359.
PR  04-OCT-2000; 2000GB-0024263.
XX
XX  (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX  Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX  WPI; 2001-488900/53.
XX
XX  Human genome-derived single exon nucleic acid probes useful for
PT  analyzing gene expression in human bone marrow -
XX
XX  Example 4; SEQ ID NO: 5438; 658bp + Sequence listing; English.
XX
XX  The present invention provides a number of single exon nucleic acid
CC  probes which are derived from genomic sequences expressed in the human
CC  bone marrow. They can be used to measure gene expression in bone marrow
CC  samples, which may enable the improved diagnosis and treatment of cancers
CC  such as lymphoma, leukemia and myeloma. The present sequence is one of
CC  the probes of the invention.
XX
XX  Sequence 472 BP; 101 A; 102 C; 101 G; 168 T; 0 other;
SQ

```

```

Query Match          3.0%; Score 44.6; DB 22; Length 472;
Best Local Similarity 65.7%; Pred. No. 0.0035;
Matches 65; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

```

```

QY  1051 GTCGATGCTGGTTCCGATTGATTGTTTCTAACACAACTTAATAGATTAAACAGCAC 1110
    || || || || || || || || || || || || || || || || || || || || || ||
DB  164 GTACTGATTTTTCAGGCTCTGTGAGATTACCAAACTTGCTGCTTAAACACAG 223
QY  1111 GGATTATTCTCACATGTTTGGACGCCGCAAAATCTGA 1149
    || || || || || || || || || || || || || || || || || || || || || ||
DB  224 AAATTATTCTCTCAGCTCTAGAGGCCAGAAATCTGA 262

```

RESULT 5

AAI36794

ID AAI36794 standard; DNA; 472 BP.

AC AAI36794;

DT 17-OCT-2001 (first entry)

DE Probe #5480 used to measure gene expression in human placenta sample.

XX Probe: microarray; human; placenta; antenatal diagnosis;

KW genetic disorder; ss.

XX Homo sapiens.

XX MO200157272-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00663.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

```

PR  21-SEP-2000; 2000US-0234687.
PR  27-SEP-2000; 2000US-0236359.
PR  04-OCT-2000; 2000GB-0024263.
XX
XX  (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX  Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX  WPI; 2001-488897/53.
XX
XX  Human genome-derived single exon nucleic acid probes useful for
PT  analyzing gene expression in human placenta -
XX
XX  Claim 25; SEQ ID NO 5480; 654bp; English.
XX
XX  The present invention relates to single exon nucleic acid probes (SENP).
CC  The present sequence is one such probe. The probes are useful for
CC  producing a microarray for predicting, measuring and displaying gene
CC  expression in samples derived from human placenta. The probes are useful
CC  for antenatal diagnosis of human genetic disorders.
XX
XX  Sequence 472 BP; 101 A; 102 C; 101 G; 168 T; 0 other;
SQ

```

```

Query Match          3.0%; Score 44.6; DB 22; Length 472;
Best Local Similarity 65.7%; Pred. No. 0.0035;
Matches 65; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

```

```

QY  1051 GTCGATGCTGGTTCCGATTGATTGTTTCTAACACAACTTAATAGATTAAACAGCAC 1110
    || || || || || || || || || || || || || || || || || || || || || ||
DB  164 GTACTGATTTTTCAGGCTCTGTGAGATTACCAAACTTGCTGCTTAAACACAG 223
QY  1111 GGATTATTCTCACATGTTTGGACGCCGCAAAATCTGA 1149
    || || || || || || || || || || || || || || || || || || || || || ||
DB  224 AAATTATTCTCTCAGCTCTAGAGGCCAGAAATCTGA 262

```

RESULT 6

ABS05629

ID ABS05629 standard; DNA; 472 BP.

AC ABS05629;

DT 19-AUG-2002 (first entry)

DE Human genome-derived single exon probe from lung SEQ ID NO 5620.

XX Human; ds: single exon probe; asthma; lung cancer; COPD; ILD;

XX chronic obstructive pulmonary disease; interstitial lung disease;

XX familial idiopathic pulmonary fibrosis; neurofibromatosis;

XX tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;

XX Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;

XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karsenger syndrome;

XX pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;

XX primary ciliary dyskinesia; pulmonary hypertension;

XX hyaline membrane disease.

XX Homo sapiens.

XX WO200186003-A2.

XX 15-NOV-2001.

XX 30-JAN-2001; 2001WO-US00665.

XX 04-FEB-2000; 2000US-180312P.

XX 26-MAY-2000; 2000US-207456P.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-234687P.

XX 27-SEP-2000; 2000US-236359P.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX 07-SEP-2001.
PD
XX
XX 28-FEB-2001; 2001WO-US06376.
PF
XX
XX 28-FEB-2000; 2000US-0515715.
PR
XX
XX (DECO-) DECODE GENETICS EHF.
PA
PI Stefansson H, Steinthorsdottir V, Gulcher JR;
XX
XX WPI: 2001-550179/61.
DR P-PSDB: AAG67900, AAG67901, AAG67902, AAG67903, AAG67904, AAG67905,
DR AAG67906, AAG67907, AAG67908, AAG67909, AAG67910, AAG67911, AAG67912,
DR AAG67913, AAG67914, AAG67915, AAG67916, AAG67917, AAG67918, AAG67919,
DR AAG67920, AAG67921, AAG67922, AAG67923, AAG67924, AAG67925, AAG67926,
DR AAG67927, AAG67928, AAG67929, AAG67930, AAG67931, AAG67932, AAG67933,
DR AAG67934, AAG67935, AAG67936, AAG67937.
XX
XX Neuiregulin-1 associated gene 1 nucleic acids and fragments, useful for
PT preventing diagnosing and treating schizophrenia -
XX
XX Disclosure: Page 90-501: 750pp; English.
XX
XX This sequence represents the human neuiregulin-1 associated gene 1
CC (NRG1AG1) of the invention. The NRG1AG1 gene is also referred to as the
CC human Schizophrenia gene. The invention also relates to fragments or
CC variants of the gene and the NRG1AG1 polypeptides they encode. The
CC NRG1AG1 nucleic acids and polypeptides may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate NRG1AG1
CC expression. For example, they may be used to treat disorders associated
CC with decreased expression by rectifying mutations or deletions in a
CC patient's genome that affect the activity of NRG1AG1 by expressing
CC inactive proteins or to supplement the patients own production of
CC NRG1AG1. Additionally, the gene may be used to produce NRG1AG1
CC polypeptides, by inserting the nucleic acids into a host cell and
CC culturing the cell to express the protein. The gene may also be used as
CC DNA probes and primers in diagnostic assays to detect and quantitate the
CC presence of similar nucleic acids in samples, and therefore which
CC patients may be in need of restorative therapy. The NRG1AG1 polypeptides
CC may also be used as antigens in the production of antibodies against
CC NRG1AG1 and in assays to identify modulators of NRG1AG1 expression and
CC activity. Anti-NRG1AG1 antibodies and antagonists may also be used to
CC down regulate expression and activity. Anti-NRG1AG1 antibodies may
CC also be used as diagnostic agents for detecting the presence of NRG1AG1
CC polypeptides in samples. NRG1AG1 is associated with schizophrenia which
CC may be prevented, diagnosed and/or treated by the above methods.
XX
SQ Sequence 1503900 BP; 452487 A; 281874 C; 288074 G; 480092 T; 1373 other;
Query Match 2.9%; Score 43.6; DB 22; Length 1503900;
Best Local Similarity 67.8%; Pred. No. 0.7;
Matches 61; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

KW Human; neuiregulin 1 gene; schizophrenia; gene therapy; ds.
XX
XX OS Homo sapiens.
XX
XX WO200164877-A2.
XX
XX
XX 07-SEP-2001.
PD
XX
XX 28-FEB-2001; 2001WO-US06377.
PF
XX
XX 28-FEB-2000; 2000US-0515716.
PR
XX
XX (DECO-) DECODE GENETICS EHF.
PA
PI Stefansson H, Steinthorsdottir V, Gulcher JR;
XX
XX WPI: 2001-514841/56.
DR P-PSDB: AAG67938, AAG67939, AAG67940, AAG67941, AAG67942, AAG67943,
DR AAG67944, AAG67945, AAG67946, AAG67947, AAG67948, AAG67949,
DR AAG67950, AAG67951, AAG67952, AAG67953, AAG67954, AAG67955,
DR AAG67956, AAG67957, AAG67958, AAG67959, AAG67960, AAG67961,
DR AAG67962, AAG67963, AAG67964, AAG67965, AAG67966, AAG67967,
DR AAG67968, AAG67969, AAG67970, AAG67971, AAG67972, AAG67973,
DR AAG67974, AAG67975.
XX
XX Neuiregulin 1 nucleic acids and proteins useful for diagnosing
PT preventing and treating schizophrenia -
XX
XX Disclosure: Page 345-756: 756pp; English.
XX
XX This sequence represents the human neuiregulin 1 gene of the invention.
CC The invention also relates to fragments or variants of the neuiregulin 1
CC gene. The gene and its proteins may be used in the prevention, diagnosis
CC and treatment of diseases associated with inappropriate neuiregulin 1
CC expression, such as schizophrenia. For example they may be used to treat
CC disorders associated with decreased neuiregulin 1 expression by rectifying
CC mutations or deletions in a patient's genome that affect the activity of
CC neuiregulin 1 by expressing inactive proteins or to supplement the
CC patients own production of polypeptides. Additionally, the gene may be
CC used to produce the neuiregulin 1 protein, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the protein. The gene
CC and its complementary sequences may also be used as DNA probes in
CC diagnostic assays to detect and quantitate the presence of similar
CC nucleic acids in samples, and therefore which patients may be in need of
CC restorative therapy. The protein may also be used as antigens in the
CC production of antibodies against neuiregulin 1 and in assays to identify
CC modulators of neuiregulin 1 expression and activity. The antibodies and
CC antagonists may also be used to down regulate expression and activity.
CC The antibodies may also be used as diagnostic agents for detecting the
CC presence of neuiregulin 1 in samples.
XX
SQ Sequence 1503900 BP; 452487 A; 281874 C; 288074 G; 480092 T; 1373 other;
Query Match 2.9%; Score 43.6; DB 22; Length 1503900;
Best Local Similarity 67.8%; Pred. No. 0.7;
Matches 61; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

XX Human foetal liver single exon nucleic acid probe #8419.
 DE
 XX
 XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
 KW
 XX
 OS Homo sapiens.
 XX
 XX WO200157277-A2.
 PN
 XX
 PD 09-AUG-2001.
 XX
 XX 30-JAN-2001; 2001WO-US00669.
 PF
 XX 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 PA
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 DR WPI: 2001-483447/52.
 XX
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human foetal liver -
 PS Claim 1; SEQ ID NO 8419; 639pp + sequence listing; English.
 XX
 CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC foetal liver. The present sequence is a single exon nucleic acid
 CC probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 585 BP; 195 A; 110 C; 131 G; 149 T; 0 other;
 Query Match 2.9%; Score 43.2; DB 22; Length 585;
 Best Local Similarity 71.2%; Pred. No. 0.011;
 Matches 57; Conservative 0; Mismatches 23; Indels 0; Gaps 0;
 QY 1081 TAAACAACCTAATAGATTAAACAGACGATTATTCACATGTTTGAGACGCCA 1140
 DB 478 TTACCCAACCTGTTGTTTAAACACACACATTTATCTCTTAGTCTGAGTCCA 419
 QY 1141 GAAATCTGACACCATTTCA 1160
 DB 418 GAACTCTGAATCAGTACA 399
 RESULT 11
 ID ABA28474/c
 ID ABA28474 standard; DNA; 585 BP.
 AC ABA28474;
 XX
 XX 23-JAN-2002 (first entry)
 DE
 XX Probe #6940 for gene expression analysis in human heart cell sample.
 XX
 KW Human; gene expression; heart; microarray; vascular system; probe;
 KW cardiovascular disease; hypertension; cardiac arrhythmia;
 KW congenital heart disease; ss.
 OS Homo sapiens.
 XX

PN WO200157274-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00666.
 XX
 XX 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 PA
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 DR WPI: 2001-488899/53.
 XX
 XX Single exon nucleic acid probes for analyzing gene expression in human
 PT hearts -
 PS Claim 1; SEQ ID NO 6940; 530pp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes for
 CC measuring human gene expression in a sample derived from human heart. The
 CC present sequence is one such probe. The probes may be used for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from the human heart via microarrays. By measuring gene expression, the
 CC probes are useful for predicting, diagnosing, grading, staging,
 CC monitoring and prognosing diseases of the human heart and vascular system
 CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
 CC congenital heart disease.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 585 BP; 195 A; 110 C; 131 G; 149 T; 0 other;
 Query Match 2.9%; Score 43.2; DB 22; Length 585;
 Best Local Similarity 71.2%; Pred. No. 0.011;
 Matches 57; Conservative 0; Mismatches 23; Indels 0; Gaps 0;
 QY 1081 TAAACAACCTAATAGATTAAACAGACGATTATTCACATGTTTGAGACGCCA 1140
 DB 478 TTACCCAACCTGTTGTTTAAACACACACATTTATCTCTTAGTCTGAGTCCA 419
 QY 1141 GAAATCTGACACCATTTCA 1160
 DB 418 GAACTCTGAATCAGTACA 399
 RESULT 12
 ID AAK08390/c
 ID AAK08390 standard; DNA; 585 BP.
 AC AAK08390;
 XX
 XX 05-NOV-2001 (first entry).
 DE
 XX Human brain expressed single exon probe SEQ ID NO: 8381.
 XX
 KW Human; brain expressed exon; gene expression analysis; probe;
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
 KW epilepsy; cancer; ss.
 OS Homo sapiens.
 XX
 XX WO200157275-A2.
 PN
 XX
 PD 09-AUG-2001.
 XX

PF 30-JAN-2001; 2001WO-US00667.
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
XX Example 4; SEQ ID NO: 8381; 650bp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX
XX Sequence 585 BP; 195 A; 110 C; 131 G; 149 T; 0 other;
SQ
Query Match 2.9%; Score 43.2; DB 22; Length 585;
Best Local Similarity 71.2%; Pred. No. 0.011; Mismatches 23; Indels 0; Gaps 0;
Matches 57; Conservative 0; Mismatches 23; Indels 0; Gaps 0;
QY 1081 TAACCAAACTTAATAGATTAAACAGCAGGATTTCTCTCAGATGTTTGAGAGCGCA 1140
DB 478 TTACCCAACTGCTGTTGTTAAACACACACATTTATCTCTTAATGTTCTGAGTCGA 419
QY 1141 GAAATCTGACACGAGTTTCA 1160
DB 418 GAACTCTGAATCAGTAACA 399
RESULT 13
AAK34270/C
ID AAK34270 standard; DNA; 585 BP.
XX
XX AAK34270;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human bone marrow expressed single exon probe SEQ ID NO: 8827.
XX
XX Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
XX Homo sapiens.
XX
XX WO200157276-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00668.
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX

PA (MOLE-) MOLECULAR DYNAMICS INC.
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
XX Example 4; SEQ ID NO: 8827; 658bp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
XX Sequence 585 BP; 195 A; 110 C; 131 G; 149 T; 0 other;
SQ
Query Match 2.9%; Score 43.2; DB 22; Length 585;
Best Local Similarity 71.2%; Pred. No. 0.011; Mismatches 23; Indels 0; Gaps 0;
Matches 57; Conservative 0; Mismatches 23; Indels 0; Gaps 0;
QY 1081 TAACCAAACTTAATAGATTAAACAGCAGGATTTCTCTCAGATGTTTGAGAGCGCA 1140
DB 478 TTACCCAACTGCTGTTGTTAAACACACACATTTATCTCTTAATGTTCTGAGTCGA 419
QY 1141 GAAATCTGACACGAGTTTCA 1160
DB 418 GAACTCTGAATCAGTAACA 399
RESULT 14
AAI39993/C
ID AAI39993 standard; DNA; 585 BP.
XX
XX AAI39993;
XX
XX 17-OCT-2001 (first entry)
XX
XX Probe #8679 used to measure gene expression in human placenta sample.
XX
XX Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200157272-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00663.
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
XX Claim 25; SEQ ID No 8679; 654bp; English.
XX

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 16, 2003, 23:50:11 ; Search time 374.434 Seconds
(without alignments)
12034.813 Million cell updates/sec

Title: US-09-605-042a-1_COPY_4000_6000
Perfect score: 2001
Sequence: 1 aagtaacctgtgtatataca.....tgatatatatatacatataca 2001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: N_Geneseq_101002:*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2001	100.0	9345	21	AAZ93966
2	63	3.1	2278	23	ABL12518
3	62.6	3.1	3001	21	AAH51803
4	62.2	3.1	8447	24	ABH80093
5	62	3.1	448	18	AAI68682
6	61.2	3.1	3032	24	ABL88369
7	60.4	3.0	240825	22	AAAF24497
8	60.2	3.0	369	21	AAH74226
9	59.8	3.0	279	21	AAH74225

C 10	59.6	3.0	382	15	AAO55246	Grapevine ribosome
C 11	59.6	3.0	1645	20	AAZ79360	Rat apoptosis inh1
C 12	59.4	3.0	14809	22	AAZ34708	Human DNA for a no
C 13	59.2	3.0	849	21	ABH81131	Shrimp polynucleot
C 14	59.2	3.0	1721	24	ABH88366	Pain regulated cDN
C 15	59.2	3.0	6727	18	AAH88014	Murine IL-5 cDNA g
C 16	59.2	3.0	6727	21	AAH73648	Murine IL-5 nucleot
C 17	59	2.9	79	13	AAO33936	Microsatellite seq
C 18	59	2.9	926	17	ABH81102	Shrimp polynucleot
C 19	59	2.9	30967	21	ABH81165	Calpain large subu
C 20	58.8	2.9	226	13	AAO27097	Shrimp polynucleot
C 21	58.4	2.9	3288	17	AAH33800	Shrimp polynucleot
C 22	58.2	2.9	3288	19	AAH27591	Erastle X diagenost
C 23	58.2	2.9	3288	20	AAH01921	Murine interleukin
C 24	58.2	2.9	3288	20	AAH01921	Murine interleukin
C 25	58.2	2.9	3288	21	AAH61238	Murine IL-17R cDNA
C 26	58.2	2.9	3288	21	AAH51987	Murine interleukin
C 27	58.2	2.9	3288	21	AAH52145	Murine interleukin
C 28	58.2	2.9	3288	21	AAH59870	Murine interleukin
C 29	58.2	2.9	3288	22	AAH02813	Murine interleukin
C 30	58.2	2.9	3288	22	AAH57186	Murine interleukin
C 31	58	2.9	316	22	AAH29391	Murine IL-17R poly
C 32	58	2.9	1131	20	AAH28294	Drosophila melanog
C 33	58	2.9	4223	19	AAV22752	Rat neutrophil imm
C 34	58	2.9	4223	19	AAV22737	Babesia microti BM
C 35	58	2.9	4223	20	AAH90001	Babesia microti BM
C 36	58	2.9	4223	20	AAH90015	Babesia microti an
C 37	58	2.9	4223	21	AAH65098	Babesia microti an
C 38	58	2.9	4223	21	AAH65098	Babesia microti an
C 39	58	2.9	4223	24	ABL89328	B. microti BMNI-7
C 40	58	2.9	4223	24	ABL89328	Babesia microti an
C 41	57.8	2.9	11617	22	AAH75374	Babesia microti an
C 42	57.8	2.9	11617	22	AAH75374	Babesia microti an
C 43	57.6	2.9	8162	24	ABK39927	Human immune/haema
C 44	57.6	2.9	18534	24	ABK84757	Human genomic DNA
C 45	57.6	2.9	26997	22	AAH46748	Human chemically p

ALIGNMENTS

AAZ93966	AAZ93966 standard; DNA; 9345 BP.
AAZ93966	AAZ93966; (first entry)
AAZ93966	29-AUG-2000 (first entry)
AAZ93966	Mouse uromodulin promoter.
AAZ93966	Uromodulin; promoter; kidney; urine; heterologous gene; treatment;
AAZ93966	therapy; gene expression; pharmaceutical; mouse; ds.
AAZ93966	Mus musculus.
AAZ93966	Location/Qualifiers
AAZ93966	4..5
AAZ93966	/tag- a
AAZ93966	/label= Polycloning site of pbs
AAZ93966	7095..7114
AAZ93966	/tag- b
AAZ93966	/label= Exon 1
AAZ93966	/cons-splice= (5'site:NO, 3'site:YES)
AAZ93966	7115..7892
AAZ93966	/tag- c
AAZ93966	/label= Intron 1
AAZ93966	7893..8073
AAZ93966	/tag- d
AAZ93966	/label= Exon 2
AAZ93966	/cons-splice= (5'site:NO, 3'site:YES)
AAZ93966	8074..8818
AAZ93966	/tag- e

FT		/label= Intron 2
FT	misc_feature	8273..8274
FT		/*tag= f
FT		/note= "APA/KPN Junction"
FT	exon	8819..9345
FT		/*tag= g
FT		/label= Exon 3
XX		
PN	WO200029608-A1.	
PD		
XX	25-MAY-2000.	
XX		
XX	12-NOV-1999;	99WO-US26870.
XX		
PR	13-NOV-1998;	98US-0108195.
PR	09-JUL-1999;	99US-0142925.
XX		
PA	(UYNV) UNIV NEW YORK STATE.	
XX		
PI	Wu X, Sun T;	
XX		
DR	WPI: 2000-387816/33.	
XX		
PT	New kidney-specific promoter useful for production of transgenic	
PT	animals as urinary bioreactors, is operably linked to a heterologous	
PT	gene	
XX		
PS	Claim 3; Fig 7a-7h; 55pp; English.	
XX		
CC	New methods to produce heterologous recombinant proteins in urine	
CC	require the use of a DNA molecule which is a kidney-specific	
CC	promoter, such as the uromodulin promoter, operably linked to a	
CC	heterologous gene encoding a biologically active protein. The	
CC	uromodulin promoter expresses the heterologous gene in vivo in the	
CC	kidneys to produce a recombinant biologically active protein in the	
CC	urine. The recombinant proteins produced may be useful for treating	
CC	human diseases. The major advantages of using this urine-based system	
CC	over milk-based systems are the ability to harvest the product soon	
CC	after birth and throughout the life of the animal irrespective of sex	
CC	or reproductive status, and the ease of product purification from	
CC	urine. In addition, livestock urine is a proven, currently utilised	
CC	source of pharmaceuticals.	
XX		
XX	Sequence 9345 BP; 2622 A; 2072 C; 2084 G; 2565 T; 2 other;	

Query Match	100.0%	Score 2001;	DB 21;	Length	9345;
Best Local Similarity	100.0%;	Pred. No. 0;			
Matches 2001; Conservative	0;	Mismatches	0;	Indels	0;
				Gaps	0;

Oy	1	AAATTAACCTCTGTTATTAACAATCCATATCAAAAGCCCTGCTGTTAGTACAAAGTTAAG	60
Db	4000	AAGTAACTCTTGTTATTAACAATCCATATCAAAAGCCCTGCTGTTAGTACAAAGTTAAG	4059
Oy	61	AAAAGATTATTCACAAGATCCAAAGTCTCCTCTTCAAAAACAAGTGTGTACAAACATTGTC	120
Db	4060	AAAAGATTATTCACAAGATCCAAAGTCTCCTCTTCAAAAACAAGTGTGTACAAACATTGTC	4119
Oy	121	TGAGGAGTAAAGTTCGATTTGGCAACATGATGTCCTTAATGGGTGAGAAATTCAGTG	180
Db	4120	TGAGGAGTAAAGTTCGATTTGGCAACATGATGTCCTTAATGGGTGAGAAATTCAGTG	4179
Oy	181	GAGTTGGCAGTCAGAAAGCACACTGCTGTAATAATGAGAGAAATAGATATATCTTTGAG	240
Db	4180	GAGTTGGCAGTCAGAAAGCACACTGCTGTAATAATGAGAGAAATAGATATATCTTTGAG	4239
Oy	241	AAATTTGGCTCAAAAAGTAGGATTAACAATTTACTTGGTGTCTGCTGATCAATTTGGGTG	300
Db	4240	AAATTTGGCTCAAAAAGTAGGATTAACAATTTACTTGGTGTCTGCTGATCAATTTGGGTG	4299
Oy	301	TCTCTGTAGCTTAGCTTACATAGAGACAGAAATAGTAGAAGGAGAGAGGAGACATT	360
Db	4300	TCTCTGTAGCTTAGCTTACATAGAGACAGAAATAGTAGAAGGAGAGAGGAGACATT	4359

OY	3 61	GGAGCACCBAAGAGAGAGAGGACCTTCCTCCCTAAAGTGAATGAAGGGGCTTCATTCGA	420
Db	4 360	GGAGGACCCBAAGAGAGAGGAGCCTTCCTCCCTAAAGTGAATGAAGGGGCTTCATTCGA	4419
OY	4 21	AGGAGAAAGGATTCAGAGTGGCCCGGGAAGATGAGGGACCAACATCCACAAGAAATGGCAG	480
Db	44 20	AGGAGAGAGATTCAGAGTGGCCCGGGAAGATGAGGGACCAACATCCACAAGAAATGGCAG	4479
OY	481	GAAGTCATCCTGTGTGCATAAATGAGAGAGGGGGGTCCAAAGATGAGACCAAGAGATGA	540
Db	4480	GAAGTCATCCTGTGTGCATAAATGAGAGAGGGGGGTCCAAAGATGAGACCAAGAGATGA	4539
OY	541	GCAAGAAAAATGGTGGATGTGGATCTCTAGAGATGGCCTGTGTGTGACCAAAATGTG	600
Db	4540	GCAAGAAAAATGGTGGATGTGGATCTAGAGATGGCCTGTGTGTGACCAAAATGTG	4599
OY	601	GGCAAAAGGCACTCCATCAACAAGACACTTCCTGTGTGAGATCCCTAAATAAAG	660
Db	4600	GGCAAAAGGCACTCCATCAACAAGACACTTCCTGTGTGAGATCCCTAAATAAAG	4659
OY	661	CACATGGCATGGCATGAGAGGCTAGGGGAGTGGAGGGGAAAGTATATAGATAGATGAGA	720
Db	4660	CACATGGCATGGCATGAGAGGCTAGGGGAGTGGAGGGGAAAGTATATAGATAGTGCAGA	4719
OY	721	AGTACAGAGAGACAGAGAAAGACAGAGATAGGAGGGACAGGTTTGCACAAAGCTTTGTC	780
Db	4720	AGTACAGAGAGACAGAGAAAGACAGAGATAGGAGGGACAGGTTTGCACAAAGCTTTGTC	4779
OY	781	CTCTCCCCACAGCTCTCTCTCCCTTCTGTATATGCACATACACAGTACACTAGTGTGCA	840
Db	4780	CTCTCCCCACAGCTCTCTCTCCCTTCTGTATATGCACATACACAGTACACTAGTGTGCA	4839
OY	841	TATGTGTGCATATGCATGTGTGTAAGACAGAGCCAGTCTGGGGTGCACGTCTTCAGGCC	900
Db	4840	TATGTGTGCATATGCATGTGTGTAAGACAGAGCCAGTCTGGGGTGTCACTTCAGGCC	4899
OY	901	CTATCTACCTGTTTTTGAGACAATCTCACTTAGTAGTAGTGAAGTCACTCCCTAGTATT	960
Db	4900	CTATCTACCTGTTTTTGAGACAATCTCACTTAGTAGTAGTGAAGTCACTCCCTAGTATT	4959
OY	961	CTACAGAGGTTTCCCAAGTGGGGAGAGAAATGGTGGGAGAGAAATTTAAAGCTGGTGG	1020
Db	4960	CTACAGAGGTTTCCCAAGTGGGGAGAGAAATGGTGGGAGAGAAATTTAAAGCTGGTGG	5019
OY	1021	ATTTCTTGAAATTCAGTGGGCTTGGGAAATAGACGCTATATATTCAGTTTCTCTGTTCT	1080
Db	5020	ATTTCTTGAAATTCAGTGGGCTTGGGAAATAGACGCTATATATTCAGTTTCTCTGTTCT	5079
OY	1081	GGCTGGCTTCCGGGGTGATCAAGACAGAGATATAGTACCCCTGTGTGGCAGTCAACACAA	1140
Db	5080	GGCTGGCTTCCGGGGTGATCAAGACAGAGATATAGTACCCCTGTGTGGCAGTCAACACAA	5139
OY	1141	GGAGACAGAAATAGGGGATGGCTCTGTGTGTGGCTGGTAGACATAGGAAAGAAATCCCTGT	1200
Db	5140	GGAGACAGAAATAGGGGATGGCTCTGTGTGTGGCTGGTAGACATAGGAAAGAAATCCCTGT	5199
OY	1201	AGCAAGATTTTGGCATTCCCAAGAGACTTAGAAGCCACAGAAAGTTTGTCTCCACAGGA	1260
Db	5200	AGCAAGATTTTGGCATTCTCCAGAGACTTAGAAGCCACAGAAAGTTTGTCTCCACAGGA	5259
OY	1261	CCAGCCAGCACTGAGAGCTGGAATGCATCAAAATCCAGAGACAGAAAGACAGGTTGCTTAGCA	1320
Db	5260	CCAGCCAGCACTGAGAGCTGGAATGCATCAAAATCCAGAGACAGAAAGACAGGTTGCTTAGCA	5319
OY	1321	CTTAGAGAGAGACACTAGCCCAAAGTCTCTCTGTCTCTGTCTTGAAGTTTGGCCAAATCTGT	1380
Db	5320	CTTAGAGAGAGACACTAGCCCAAAGTCTCTCTGTCTCTGTCTTGAAGTTTGGCCAAATCTGT	5379
OY	1381	CAAACTCTGAAAAAATAGCATCTTTTAAATTCAGAAAGGATACAAAGAGACAACTTACATG	1440
Db	5380	CAAACTCTGAAAAAATAGCATCTTTTAAATTCAGAAAGGATACAAAGAGACAACTTACATG	5439
OY	1441	GGACCTTTGAAAAAAGCATAGGGCATCAATACTAAAGTTACAAAGATAACAATCACTGCG	1500

```

Db 5440 GGAACCTGTAAAGAACATGAGCATGATCACTAAAGTTACAAAGATTAACAAATCAGTGG 5499
    |||||||
Qy 1501 TGAGTGAACAAAGACATGGCCATGTTTTTTTGTATGAACACACAGCAGCAGCAGG 1560
    |||||||
Db 5500 TGAGTGAACAAAGACATGGCCATGTTTTTTTGTATGAACACACAGCAGCAGCAGG 5559
    |||||||
Qy 1561 CACTGACGTGTGCGCAGCGCGCAGACACACAGCGGACACACACACACACATGCA 1620
    |||||||
Db 5560 CACTGACGTGTGCGCAGCGCGCAGACACACAGCGGACACACACACACACATGCA 5619
    |||||||
Qy 1621 CACATGACACACACAACTGCAAAAGTGAATTAATTAATTTCTCAGCTTTGGCAAG 1680
    |||||||
Db 5620 CACATGACACACACAACTGCAAAAGTGAATTAATTAATTTCTCAGCTTTGGCAAG 5679
    |||||||
Qy 1681 TGAGTGAAGTGTGATCAAAAGTGAATTAATTAATTTCTCAGCTTTGGAGAGG 1740
    |||||||
Db 5680 TGAGTGAAGTGTGATCAAAAGTGAATTAATTAATTTCTCAGCTTTGGAGAGG 5739
    |||||||
Qy 1741 TTATTAATTTATTTGTTATTTATTTCTATTTTCTGTTGTTGCGACGCTTAAGTTGCT 1800
    |||||||
Db 5740 TTATTAATTTATTTGTTATTTATTTCTATTTTCTGTTGTTGCGACGCTTAAGTTGCT 5799
    |||||||
Qy 1801 TGAACCTGATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1860
    |||||||
Db 5800 TGAACCTGATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 5859
    |||||||
Qy 1861 AGTGGCCAGATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1920
    |||||||
Db 5860 AGTGGCCAGATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 5919
    |||||||
Qy 1921 CCAATTATTAACAACACACACATATATACACATATACACACACACACACATATATAT 1980
    |||||||
Db 5920 CCAATTATTAACAACACACACATATATACACATATATACACACACACACACATATATAT 5979
    |||||||
Qy 1981 ATGTATATATATATATATATATATATATATATATATATATATATATATATATATAT 2001
    |||||||
Db 5980 ATGTATATATATATATATATATATATATATATATATATATATATATATATATATAT 6000
    |||||||

```

RESULT 2

ABLI2518

ID ABLI2518 standard; cDNA; 2278 BP.

XX ABLI2518;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 32036.

KW Drosophila; developmental biology; cell signalling; insecticide;

OS Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI: 2001-656860/75.

XX P-PSDB; ABB68415.

PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell

PT Interactions -

XX Claim 1; SEQ ID NO 32036; 21bp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
 CC sequences (AB101840-AB161175) and the encoded proteins
 CC (AB57737-AB572072).

CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 2278 BP; 694 A; 498 C; 419 G; 667 T; 0 other;

Query Match 3.1%; Score 63; DB 23; Length 2278;

Best Local Similarity 66.7%; Pred. No. 2.7e-07; Mismatches 45; Indels 0; Gaps 0;

Matches 90; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

```

Qy 1540 AATACACCCACAGGACAGGACATGCTGCGCAGCGGCGACACACACACCGCC 1599
    |||||||
Db 1357 ACACACACACACACACACACACACACACACACACACACACACACACACACACAC 1416
    |||||||
Qy 1600 ACACACACACACACCGCATGACATGACACACACACACAACTGCAAAAGTGAATTAAG 1659
    |||||||
Db 1417 ACACACACACACACACACACACACACACACACACACACACACACACACACACAC 1476
    |||||||
Qy 1660 AATATTCACCTTGG 1674
    |||||||
Db 1477 AATTTCAATTAATTG 1491
    |||||||

```

RESULT 3

AAH51803/c

ID AAH51803 standard; DNA; 3001 BP.

XX AAH51803;

XX 29-AUG-2001 (first entry)

DE Chromosome 13q31-q33 bi-allelic marker containing amplicon SEQ ID 215.

KW sbg1; g34665; sbg2; g35017; g35018; chromosome 13q31-q33; haplotype;
 KW bi-allelic marker; polymorphism; schizophrenia; bipolar disorder; ds.

XX Homo sapiens.

XX WO2000058510-A2.

XX 05-OCT-2000.

XX 30-MAR-2000; 2000WO-IB00435.

XX 30-MAR-1999; 99US-0126903.

XX 30-APR-1999; 98US-0131971.

XX 30-APR-1999; 99US-0132065.

XX 14-JUL-1999; 99US-0143928.

XX 29-JUL-1999; 99US-0145915.

XX 29-JUL-1999; 99US-0146452.

XX 28-OCT-1999; 99US-0146453.

XX 28-OCT-1999; 99US-0162288.

XX (GENSET) GENSET.
 XX Cohen D, Blumenfeld M, Chumakov I, Bougueteloret L, Bihain B;
 XX Estioux L;
 XX WPI: 2000-619082/59.
 XX Polynucleotides comprising sequences from sbg1 and g35018 bi-allelic

PT markers are used for genotyping and detecting schizophrenia or bipolar
PT disorder and predisposition to these disorders -

PS Claim 2: Page 716-717: 737pp: English.

CC AAH51601 represents a human genomic nucleotide sequence comprising sbg1,
CC g34665, sbg2, g35017 and g35018 nucleic acid sequences located on the
CC human chromosome 13q31-q33 locus. The nucleotide sequences contain
CC diallelic markers and polymorphisms. Sequences AAH51602 - AAH51626 and
CC AAH62907 - AAH62915 represent cDNA human sbg1 cDNA sequences and protein
CC products. AAH51627 - AAH51631 and AAH62916 - AAH62918 represent g35018
CC cDNA sequences and protein products. Primers AAH51632 - AAH51699 are used
CC to isolate sbg1 cDNAs, while sbg1 exons from different primates are
CC represented by sequences AAH51642 - AAH51699. Nucleotide sequences of
CC amplicons which comprise diallelic markers located on the chromosome
CC 13q31-q33 locus are represented in AAH51700 - AAH51817. Diallelic markers
CC are represented in the sequences by degenerate/undefined base codes. PCR
CC primers AAH51818 and AAH51819 are used in the isolation of sequences of
CC the invention. The diallelic marker containing nucleotide sequences are
CC used to determine the identity of the nucleotide at a diallelic marker in
CC a sample DNA sequence. The nucleotide sequences may be labelled and used
CC for genotyping by determining the identity of a nucleotide at a region
CC D-related diallelic marker in a biological sample from single or multiple
CC subjects. By determining the frequency of a diallelic marker in a
CC population an association between a genotype and a trait, a haplotype and
CC a trait and a phenotype and a trait can be detected. The sequences can be
CC used to determine a predisposition to or early onset of schizophrenia or
CC bipolar disorder or a beneficial response to or side effects related to
CC treatment against schizophrenia or bipolar disorder.

SO Sequence 3001 BP: 990 A: 470 C: 494 G: 1046 T: 1 other:

Query Match 3.1%; Score 62.6; DB 21; Length 3001;
Best Local Similarity 83.5%; Pred. No. 4e-07;
Matches 71; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

OY 1917 AAGCCAAATTATACAAACACACATATATATACACATATACACACACACACATAT 1976

DB 754 AAACATAT 695

OY 1977 AT 2001

DB 694 AT 670

RESULT 4

ABN80093/C

ID ABN80093 standard; DNA: 8447 BP.

AC ABN80093;

DT 15-JUL-2002 (first entry)

DE Human chemically modified disease associated gene SEQ ID NO 110.

XX Human: development; homeobox gene; HOX; diabetes; cancer; apoptosis;
KW heart disease; epilepsy; histone deacetylation; muscular dystrophy;
KW dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;
KW antidiabetic; cyostatic; anticonvulsant; ds.

OS Homo sapiens.
OS Synthetic.

PN WO200200927-A2.

PD 03-JAN-2002.

PF 02-JUL-2001; 2001WO-EP07536.

PR 30-JUN-2000; 2000DE-103529.

PR 01-SEP-2000; 2000DE-1043826.

PA (EPig-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;
PI WPI, 2002-130908/17.

PT Novel nucleic acid useful for diagnosis and therapy of diseases
PT associated with development genes such as diabetes, comprises a
PT sequence of a segment of chemically pretreated DNA of genes associated
PT with development -

PS Claim 1: SEQ ID NO 110; 27pp: English.

CC The invention relates to a nucleic acid (I) comprising a sequence at
CC least 18 bases in length of a segment of chemically pretreated DNA (II)
CC of genes associated with development selected from 87 genes listed in
CC the specification such as ACCPN, ADFN, or APL1 and comprising one of 350
CC sequences (ABN79984-ABN80333) or their complements. The invention is
CC useful for the diagnosis or therapy of diseases associated with
CC development genes, in particular disease related to homeobox containing
CC genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes
CC associated with congenital heart disease, epilepsy, diseases related to
CC histone deacetylation, Curriarino syndrome, diseases related with the
CC development of the brain and limb girdle muscular dystrophy and dwarfism.
CC Oligomers specific to each of the genes are useful for detecting the
CC methylation state of all CpG dinucleotides within the 350 sequences or
CC (II) and their complementary sequences, as primer oligonucleotides for
CC the amplification of the 350 sequences, (II) and/or their complements and
CC as oligomer probes for detecting the cytosine methylation state and/or
CC single nucleotide polymorphisms (SNPs).

CC Note: The sequence data for this patent did not form part of the printed
CC specification but is based on sequence information supplied to Derwent by
CC the European Patent Office.

SO Sequence 8447 BP: 1957 A: 144 C: 2021 G: 4325 T: 0 other:

Query Match 3.1%; Score 62.2; DB 24; Length 8447;
Best Local Similarity 65.5%; Pred. No. 9.3e-07;
Matches 91; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

OY 1540 AAACACACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1599

DB 4122 AC 4063

OY 1600 ACACACACACACACGATCGACATCGACACACACACACAACTGCAAAAGCAATAAAG 1659

DB 4062 ACAA 4003

OY 1660 ATATTTCACCTTGGCAA 1678

DB 4002 ATTAACACTCTTTAAAAA 3984

RESULT 5

AAT68682

ID AAT68682 standard; DNA: 448 BP.

AC AAT68682;

DT 24-FEB-1998 (first entry)

DE Pennaeus vannamei genomic DNA marker insert from plasmid pV12.

XX Pennaeus vannamei; marker; satellite; microsatellite; shrimp; repeat;
KW aquaculture; reproduction; growth; size; disease resistance; ds.

OS Pennaeus vannamei.

FT Key Location/Qualifiers

FT Satellite 246..408

FT repeat_unit /note="Satellite region containing CA repeat units"

FT /tag="b

XX PN MO9721835-A2.
 XX PD 19-JUN-1997.
 XX PF 12-DEC-1996; 96MO-US19568.
 XX PR 12-DEC-1995; 95US-0570751.
 XX PA (WORC-) WORCESTER POLYTECHNIC INST.
 XX PI Bagshaw JC, Buckholt MA;
 XX DR WPI: 1997-332803/30.
 XX PT Markers containing nucleotide repeats for selecting Penaeus shrimps
 XX PT - used to select for characteristics advantageous in aquaculture
 XX PS Claim 2: Page 21; 26pp; English.
 XX CC The present sequence represents a specifically claimed genomic sequence
 CC of Penaeus vannamei shrimps which can be used as a marker to select
 CC Penaeus shrimp having a predetermined genetically transmitted
 CC characteristic. The marker comprises a nucleotide repeat present in the
 CC shrimp's genome. The presence and pattern of the marker are used to
 CC characterise the shrimps, specifically for selection (for commercial
 CC farming) of favourable growth features, e.g. increased reproduction,
 CC growth rate and size, better disease resistance, growth in cold water
 CC and improved aquaculture production. Species, strains and individuals
 CC can also be identified.
 XX SQ Sequence 448 BP; 169 A; 130 C; 41 G; 108 T; 0 other;
 Query Match 3.1%; Score 62; DB 18; Length 448;
 Best Local Similarity 68.3%; Pred. No. 2.1e-07;
 Matches 86; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
 QY 1540 AACACAGCAGCAGGACGACGACTCAGTGTGGCAGCGCGACACACACACGCGC 1599
 Db 301 AC 360
 QY 1600 AACACACACACACGATGACACATGACACACACAACTGCAAAAGTAAATAAAG 1659
 Db 361 ACGGAAGATATA 420
 QY 1660 ATATT 1665
 Db 421 ACTCT 426
 RESULT 6
 ABL88369
 ID ABL88369 standard; cDNA; 3032 BP.
 XX ABL88369;
 AC
 XX 16-MAY-2002 (first entry)
 DT
 XX Pain regulated cDNA sequence 12.
 DE
 XX Pain; analgesic; gene therapy; neurological disorder;
 KW neurodegenerative disease; gene; ss.
 XX
 OS Rattus sp..
 XX
 FN WO200212338-A2.
 PD 14-FEB-2002.
 PD 03-AUG-2001; 2001WO-EP09011.
 PF 03-AUG-2000; 2000DE-1037759.
 PR
 XX

PA (CHEF) GROENENTHAL GMBH.
 XX
 XX Gillen C, Wetzel I, Wnendt S, Weihe E, Schaefer MK;
 XX
 XX WPI: 2002-257469/30.
 DR
 XX Identifying pain-regulating compounds, useful for treating chronic pain
 PT and for diagnosis, by measuring binding of compounds to specific
 PT peptides and proteins
 XX
 XX Claim 14; Fig 7; 213pp; German.
 PS
 XX The invention relates to identifying pain-regulating substances (A)
 CC comprises (i) incubating a test substance with a cell (or preparation
 CC from it) that has synthesised a peptide or protein (B) and (ii) measuring
 CC either binding of the test substance to (B) or some functional parameter
 CC that is altered by this binding. The method is useful for identifying
 CC pain-regulating substances (A) with analgesic activity. (A) along with
 CC nucleic acid (ABL88411-ABL88441) that encode proteins (B).
 CC ABB85006-ABB85037) that interact with (A); (B); vectors containing the
 CC nucleic acid; antibodies against (B); cells that express (B) and agents
 CC that bind to (B), are all useful for treating pain, particularly chronic
 CC pain, including use in gene therapy. The same materials can also be used
 CC for diagnosis, e.g. of neurological and neurodegenerative diseases. The
 CC present sequence is that of a polynucleotide of the invention.
 XX
 SQ Sequence 3032 BP; 986 A; 615 C; 586 G; 841 T; 4 other;
 Query Match 3.1%; Score 61.2; DB 24; Length 3032;
 Best Local Similarity 56.4%; Pred. No. 1e-06;
 Matches 114; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 1420 TACAGAGAGACTTACATGAGGACCTTGTAAAGCATAGGCGATCAGTAACACT 1479
 Db 551 TAAAGTTTAAATCAATGAGAGACAAATTTCAAAAGATGACAGAACTGAGAA 610
 QY 1480 TACAAAGATTAACATCATGAGTGTGAGTGAACAAAGACATGCGCATGTTTGTATG 1539
 Db 611 AAAACCCAGAGATTTTCGTAACTTCATATGATACATGATCTACTCTGAGAA 670
 QY 1540 AACACAGCAGCAGGACGACGACTCAGTGTGGCAGCGCGACACACACACGCGC 1599
 Db 671 ATACACATGACAGGACGACGAGCAAGCGCGCGCGCGACACACACACACACAC 730
 QY 1600 ACACACACACACGACGATGCAC 1621
 Db 731 ACACACACACACACACACACAC 752
 RESULT 7
 AAF24497/c
 ID AAF24497 standard; cDNA; 240825 BP.
 XX AAF24497;
 AC
 XX 23-MAY-2001 (first entry)
 DT
 XX Human PG-3 gene.
 DE
 XX Human; PG-3; cancer; BRCA1; chromosome 8p23; ds.
 KW Homo sapiens.
 XX
 OS
 XX
 FN Key Location/Qualifiers
 PD misc-feature 1..2000
 PD /tag= a
 FT /note= "5' regulatory region"
 FT primer_bind 1823..1840
 FT /tag= b
 FT primer_bind 1980..1998
 FT /tag= c
 FT misc_binding 1987..2011
 FT /tag= d

```

FT      /note= "binds probe"
FT      replace(1999,C)
FT      /tag= e
FT      primer_bind
FT      complement(2000..2018)
FT      /tag= f
FT      CDS
FT      2001..238825
FT      /tag= g
FT      /product= "Pg-3"
FT      /note= "this sequence contains introns"
FT      2001..2079
FT      exon
FT      /tag= h
FT      /label= "A"
FT      primer_bind
FT      2108..2125
FT      /tag= i
FT      primer_bind
FT      4559..4577
FT      /tag= j
FT      primer_bind
FT      4582..4600
FT      /tag= k
FT      misc_binding
FT      4589..4613
FT      /tag= l
FT      /note= "binds probe"
FT      replace(4601,G)
FT      /tag= m
FT      primer_bind
FT      complement(4602..4620)
FT      /tag= n
FT      4627..4718
FT      exon
FT      /tag= o
FT      /label= "B"
FT      primer_bind
FT      4891..4908
FT      /tag= p
FT      primer_bind
FT      10007..10025
FT      /tag= q
FT      exon
FT      10115..10233
FT      /tag= r
FT      /label= "C"
FT      primer_bind
FT      10209..10227
FT      /tag= s
FT      misc_binding
FT      10216..10240
FT      /tag= t
FT      /note= "binds probe"
FT      replace(10228,T)
FT      allele
FT      /tag=
FT      primer_bind
FT      complement(10229..10247)
FT      /tag= v
FT      primer_bind
FT      10267..10285
FT      /tag= w
FT      misc_binding
FT      10274..10298
FT      /tag= x
FT      /note= "binds probe"
FT      replace(10286,T)
FT      allele
FT      /tag= y
FT      primer_bind
FT      complement(10287..10305)
FT      /tag= z
FT      replace(10370,)
FT      allele
FT      /tag= aa
FT      primer_bind
FT      10411..10430
FT      /tag= ab
FT      exon
FT      26810..26897
FT      /tag= ac
FT      /label= "D"
FT      exon
FT      31357..31471
FT      /tag= ad
FT      /label= "E"
FT      exon
FT      34261..34404
FT      /tag= ae
FT      /label= "F"
FT      exon
FT      37377..37466
FT      /tag= af
FT      /label= "S"
FT      exon
FT      39704..40858
FT      /tag= ag
FT      /label= "T"
FT      primer_bind
FT      39556..39574

FT      /tag= ah
FT      primer_bind
FT      39877..39896
FT      /tag= ai
FT      primer_bind
FT      39925..39943
FT      /tag= aj
FT      misc_binding
FT      39932..39956
FT      /tag= ak
FT      /note= "binds probe"
FT      replace(39944,T)
FT      allele
FT      /tag= al
FT      primer_bind
FT      complement(39945..39963)
FT      /tag= am
FT      primer_bind
FT      39953..39970
FT      /tag= an
FT      primer_bind
FT      39954..39972
FT      /tag= ao
FT      misc_binding
FT      39961..39985
FT      /tag= ap
FT      /note= "binds probe"
FT      replace(39973,C)
FT      allele
FT      /tag= ap
FT      primer_bind
FT      complement(39974..39992)
FT      /tag= ar
FT      40242..40259
FT      /tag= as
FT      primer_bind
FT      41137..41154
FT      /tag= at
FT      primer_bind
FT      41366..41384
FT      /tag= au
FT      misc_binding
FT      41373..71397
FT      /tag= av
FT      /note= "binds probe"
FT      replace(41385,C)
FT      allele
FT      /tag= aw
FT      41385..41403
FT      /tag= ax
FT      primer_bind
FT      complement(41386..41404)
FT      /tag= ay
FT      41392..41416
FT      /tag= az
FT      /note= "binds probe"
FT      replace(41404,C)
FT      allele
FT      /tag= ba
FT      primer_bind
FT      complement(41405..41423)
FT      /tag= bb
FT      41564..41581
FT      /tag= bc
FT      primer_bind
FT      42122..42141
FT      /tag= bd
FT      primer_bind
FT      42213..42231
FT      /tag= be
FT      42220..42244
FT      /tag= bf
FT      /note= "binds probe"
FT      replace(42232,C)
FT      allele
FT      /tag= bg
FT      primer_bind
FT      complement(42233..42251)
FT      /tag= bh
FT      42526..42543
FT      /tag= bi
FT      exon
FT      50436..50545
FT      /tag= bj
FT      /label= "G"
FT      primer_bind
FT      67289..67309
FT      /tag= bk
FT      primer_bind
FT      67456..67474
FT      /tag= bl
FT      misc_binding
FT      67463..67487
FT      /tag= bm
FT      /note= "binds probe"
FT      replace(67475,G)
FT      allele
FT      /tag= bn
FT      primer_bind
FT      complement(67476..67494)

```

```

FT primer_bind /tag= bp
FT 67724..67741
FT primer_bind /tag= bp
FT 69182..69200
FT primer_bind /tag= bq
FT 69502..69520
FT misc_binding /tag= br
FT 69509..69533
FT /tag= bs
FT /note= "binds probe"
FT replace(69521,G)
FT allele
FT /tag= bt
FT complement(69522..69540)
FT primer_bind /tag= bu
FT 69609..69626
FT primer_bind /tag= bv
FT 72698..72715
FT primer_bind /tag= bw
FT 72819..72837
FT primer_bind /tag= bx
FT 72826..72850
FT misc_binding /tag= by
FT /note= "binds probe"
FT replace(72838,T)
FT allele /tag= bz
FT complement(72839..72857)
FT primer_bind /tag= ca
FT 72881..72918
FT exon /tag= cb
FT /tag= "H"
FT /label= "H"
FT 73099..73117
FT primer_bind /tag= cc

Query Match 3.0%; Score 60.4; DB 22; Length 240825;
Best Local Similarity 74.5%; Pred. No. 1.9e-05;
Matches 76; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 1525 GTTTTGTATGTAACACACACACACACACGACTGCTGTGCGCAGCGCGCA 1584
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 234684 GATGTTGAGTGAAGAACACACACACACACACGACACACACACGCGCGCA 234625
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1585 CACACACACACGCGCACACACACACACACGATGCACACATG 1626
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 234624 CACACACACACACACACACACACACACACACACACACG 234583
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
AAA/4226/C
ID AAA/4226 standard; DNA; 369 BP.
XX
AC AAA/4226;
XX
DT 29-NOV-2000 (first entry)
XX
DE Lobloolly pine SSR locus RIPPT106.
XX
KM Lobloolly pine: Simple Sequence Repeat; SSR; microsatellite DNA repeat;
KW genetic marker; mapping; inheritance study; population genetics study;
XX plant breeding programme; ss.
XX
XX Pinus taeda.
OS
XX
XX WO200042210-A2.
PN
XX
XX 20-JUL-2000.
PD
XX
XX 06-JAN-2000; 2000WO-US00325.
PE
XX
XX 15-JAN-1999; 99US-0232884.
PR
XX 19-JAN-1999; 99US-0232785.
PA
XX (INTO ) INT PAPER CO.
PA (ECHT/) ECHT C S.
PA

```

```

PA (NELS/) NELSON C D.
PA (USDA ) US SEC OF AGRIC.
XX
XX ECHT CS, Nelson CD;
XX
XX WPI; 2000-482836/42.
DR
XX
XX Polynucleotide having simple sequence repeat useful as markers in
PT plants for genetic characterization e.g. genetic mapping study; an
PT inheritance study of a commercially important trait in a plant breeding
PT program
XX
XX Claim 1; Page 30; 57pp; English.
PS
XX
XX The present sequence is a lobloolly pine simple sequence repeat (SSR)
CC locus. SSRs are also known as microsatellite DNA repeats. The present
CC sequence is useful as a genetic marker for genetic mapping, population
CC genetics studies and inheritance studies in various plant breeding
CC programmes.
XX
XX Sequence 369 BP; 97 A; 67 C; 99 G; 106 T; 0 other;
SQ

Query Match 3.0%; Score 60.2; DB 21; Length 369;
Best Local Similarity 70.8%; Pred. No. 6.3e-07;
Matches 80; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 1887 CTATACGCTTATGCTGTGCTAAGACTAAGCCCAATTATACAAACACACATAT 1946
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 293 CTATACGCTGATACACATGACATGCTATATGCTATTCATTATACACACACACAC 234

QY 1947 ATACACATACACACACACACACACGATATATATATATATATATATACATA 1999
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 233 GCACACACACACACACACACACATATATATATATATATATATATATATATA 181
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
AAA/4225/C
ID AAA/4225 standard; DNA; 279 BP.
XX
AC AAA/4225;
XX
DT 29-NOV-2000 (first entry)
XX
DE Lobloolly pine SSR locus RIPPT104.
XX
KM Lobloolly pine: Simple Sequence Repeat; SSR; microsatellite DNA repeat;
KW genetic marker; mapping; inheritance study; population genetics study;
XX plant breeding programme; ss.
XX
XX Pinus taeda.
OS
XX
XX WO200042210-A2.
PN
XX
XX 20-JUL-2000.
PD
XX
XX 06-JAN-2000; 2000WO-US00325.
PE
XX
XX 15-JAN-1999; 99US-0232884.
PR
XX 19-JAN-1999; 99US-0232785.
PA
XX (INTO ) INT PAPER CO.
PA (ECHT/) ECHT C S.
PA (NELS/) NELSON C D.
PA (USDA ) US SEC OF AGRIC.
XX
XX ECHT CS, Nelson CD;
XX
XX WPI; 2000-482836/42.
DR
XX
XX Polynucleotide having simple sequence repeat useful as markers in
PT plants for genetic characterization e.g. genetic mapping study; an
PT inheritance study of a commercially important trait in a plant breeding
PT program

```

XX Claim 1; Page 30; 57pp; English.

PS The present sequence is a lobolly pine Simple Sequence Repeat (SSR)

CC locus. SSRs are also known as microsatellite DNA repeats. The present

CC sequence is useful as a genetic marker for genetic mapping, population

CC genetics studies and inheritance studies in various plant breeding

CC programmes.

XX Sequence 279 BP; 46 A; 26 C; 83 G; 122 T; 2 other;

SQ

Query Match 3.0%; Score 59.8; DB 21; Length 279;

Best Local Similarity 68.9%; Pred. No. 7e-07;

Matches 82; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 1540 AACACACGACAGGACGACGACTGACGTGCGACGCGGACACACACACACGCGC 1599

DB 206 AC 147

QY 1600 ACACACACACACACGATGACATGACACACACACACAAACGCAAAATGATATAAA 1658

DB 146 ACACACACACACACACACACACACACACATATCACAATTTATACGCAAAATGAA 88

RESULT 10

AA055246/C

ID AA055246 standard; DNA; 382 BP.

AC AA055246;

XX 14-JUL-1994 (first entry)

DE Grapevine ribosomal clone INVGT19C.

XX Ribosome; grapevine; Vitis; IGS region; rDNA; polymorphism;

KW grape cultivar; probe; primer; detection; ds.

XX Vitis vinifera.

OS

XX Key Location/Qualifiers

FT primer_bind 76..100

FT primer_bind complement (295..317)

FT primer_bind /*tag- a

FT /*tag- b

XX MO9401580-A.

XX 20-JAN-1994.

XX 30-JUN-1993; 93WO-AUD00320.

XX 03-JUL-1992; 92AU-0003330.

XX (CSIR) COMMONWEALTH SCI & IND RES ORG.

XX Scott NS, Thomas MR;

XX WPI; 1994-035083/04.

XX Novel ribosome DNA probe sequences - for the accurate identification

PT of grape cultivars

PS Claim 26; Fig 5p; 55pp; English.

XX The sequences given in AA055231-50 are derived from ribosomes of the

CC grapevine genus Vitis. These sequences represent the IGS region of

CC the ribosomal (r)DNA repeat and contain polymorphisms. These

CC polymorphisms may be used in a method for the identification of

CC different grape cultivars. These clones contain simple repeat

CC sequences and were identified in a genomic library of grapevine DNA

CC using simple di-, tri- or tetra- nucleotide repeats such as (AT)⁸,

CC (GT)¹⁰, (CG)¹⁰ and such like as probes.

SQ Sequence 382 BP; 122 A; 28 C; 95 G; 137 T; 0 other;

Query Match 3.0%; Score 59.6; DB 15; Length 382;

Best Local Similarity 82.9%; Pred. No. 9.5e-07;

Matches 68; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1920 CCCATTTATATCAACACACACATATATACACATATACACACACACACAGTATATA 1979

DB 195 CATATATA 136

QY 1980 TATGTATATATATATATATACATA 2001

DB 135 TATATATATATATATATATATATA 114

RESULT 11

AAAT9360/C

ID AAAT9360 standard; DNA; 1645 BP.

AC AAAT9360;

XX 14-SEP-1999 (first entry)

DE Rat apoptosis inhibitory factor PCRF35 gene.

XX Rat apoptosis inhibitory factor PCRF35 gene.

KW Neurotropic; neuroprotective; apoptosis inhibitory factor; PCRF35; rat;

KW pheochromocytoma cell; PC12 cell; human; bc112; nerve cell death;

KW Parkinson's disease; Alzheimer's disease; ss.

XX Rattus sp.

OS

XX WO9931237-A1.

XX 24-JUN-1999.

XX 11-DEC-1998; 98WO-JP05609.

XX 14-MAY-1998; 98JP-0131634.

XX 12-DEC-1997; 97JP-0343112.

XX (TAIS) TAISHO PHARM CO LTD.

XX Ohsawa Y, Uchiyama Y;

XX WPI; 1999-395180/33.

XX P-PSDB; AAY14570.

XX New protein obtained by culturing pheochromocytoma transformed by a

PT bc12 gene as a protooncogene, useful in remedies for

PT apoptosis-caused diseases. e.g. Alzheimer disease

XX Example 2; Page 39-43; 50pp; Japanese.

XX This sequence represents the gene encoding the rat apoptosis inhibitory

CC factor PCRF35. The sequence was isolated from rat pheochromocytoma PC12

CC cells transformed with the human bc12 gene. The protein can be used for

CC the treatment of apoptosis-caused diseases, including those due to nerve

CC cell death, e.g. Parkinson's disease and Alzheimer's disease.

XX

SQ Sequence 1645 BP; 359 A; 472 C; 459 G; 355 T; 0 other;

Query Match 3.0%; Score 59.6; DB 20; Length 1645;

Best Local Similarity 62.6%; Pred. No. 2.1e-06;

Matches 109; Conservative 0; Mismatches 64; Indels 1; Gaps 1;

QY 1751 TTGTTATGTTATATATATATTTTCTGTTGGACGCTAGTGGCTTGAATCACT 1810

DB 1644 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1585

QY 1811 ATGAGCTAGACAGACCTTGAGCTTCTGATCCTTATATACACTCT-CAAGTCCGAG 1869

DB 1584 GTTATACCAAGATGACTGACTGACTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1525

OY 1870 ATTATAGTGTGCACCACTACTAGTTATGCTGTGCTAGAGCACTAGCCCA 1923
DB 1524 ATACAGCGCTGAGTCACACATGCTGTTTATGTGCTGGGACAGACGCCA 1471

RESULT 12

AAS34708
ID AAS34708 standard; DNA; 14809 BP.

XX AAS34708;

DT 17-DEC-2001 (first entry)

XX Human DNA for a novel foetal antigen, SEQ ID No 2132.

XX Human: foetal tissue antigen; ds; antiinflammatory; neuroprotective;
KW immunomodulator; cardiovascular; cytosolic; nephrotoxic;
KW cardiovascular; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; breast neoplasm; cancer;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischemia; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; epithelial cell proliferation; food additive.

XX Homo sapiens.

PN W020015312-A2.

PD 02-AUG-2001.

PF 17-JAN-2001; 2001WO-US01321.

PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0186874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226686.

PR 22-AUG-2000; 2000US-0227182.

PR 23-AUG-2000; 2000US-0227009.

PR 30-AUG-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.

PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.

PR 05-SEP-2000; 2000US-0229509.

PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.

PR 08-SEP-2000; 2000US-0231242.

PR 08-SEP-2000; 2000US-0231243.

PR 08-SEP-2000; 2000US-0231244.

PR 08-SEP-2000; 2000US-0231413.

PR 08-SEP-2000; 2000US-0231414.

PR 08-SEP-2000; 2000US-0232080.

PR 08-SEP-2000; 2000US-0232081.

PR 12-SEP-2000; 2000US-0231968.

PR 14-SEP-2000; 2000US-0232397.

PR 14-SEP-2000; 2000US-0232398.

PR 14-SEP-2000; 2000US-0232399.

PR 14-SEP-2000; 2000US-0232400.

PR 14-SEP-2000; 2000US-0232401.

PR 14-SEP-2000; 2000US-0233063.

PR 14-SEP-2000; 2000US-0233064.

PR 14-SEP-2000; 2000US-0233065.

PR 21-SEP-2000; 2000US-0234223.

PR 21-SEP-2000; 2000US-0234274.

PR 25-SEP-2000; 2000US-0234997.

PR 25-SEP-2000; 2000US-0234998.

PR 26-SEP-2000; 2000US-0235484.

PR 27-SEP-2000; 2000US-0235834.

PR 27-SEP-2000; 2000US-0235836.

PR 29-SEP-2000; 2000US-0236327.

PR 29-SEP-2000; 2000US-0236367.

PR 29-SEP-2000; 2000US-0236368.

PR 29-SEP-2000; 2000US-0236369.

PR 29-SEP-2000; 2000US-0236370.

PR 02-OCT-2000; 2000US-0236802.

PR 02-OCT-2000; 2000US-0237037.

PR 02-OCT-2000; 2000US-0237038.

PR 02-OCT-2000; 2000US-0237039.

PR 02-OCT-2000; 2000US-0237040.

PR 13-OCT-2000; 2000US-0239935.

PR 13-OCT-2000; 2000US-0239937.

PR 20-OCT-2000; 2000US-0240960.

PR 20-OCT-2000; 2000US-0241221.

PR 20-OCT-2000; 2000US-0241785.

PR 20-OCT-2000; 2000US-0241786.

PR 20-OCT-2000; 2000US-0241787.

PR 20-OCT-2000; 2000US-0241808.

PR 20-OCT-2000; 2000US-0241809.

PR 20-OCT-2000; 2000US-0241826.

PR 01-NOV-2000; 2000US-0244617.

PR 08-NOV-2000; 2000US-0246474.

PR 08-NOV-2000; 2000US-0246475.

PR 08-NOV-2000; 2000US-0246476.

PR 08-NOV-2000; 2000US-0246477.

PR 08-NOV-2000; 2000US-0246478.

PR 08-NOV-2000; 2000US-0246523.

PR 08-NOV-2000; 2000US-0246524.

PR 08-NOV-2000; 2000US-0246525.

PR 08-NOV-2000; 2000US-0246526.

PR 08-NOV-2000; 2000US-0246527.

PR 08-NOV-2000; 2000US-0246528.

PR 08-NOV-2000; 2000US-0246532.

PR 08-NOV-2000; 2000US-0246609.

PR 08-NOV-2000; 2000US-0246610.

PR 08-NOV-2000; 2000US-0246611.

PR 08-NOV-2000; 2000US-0246613.

PR 17-NOV-2000; 2000US-0249207.

PR 17-NOV-2000; 2000US-0249208.

PR 17-NOV-2000; 2000US-0249209.

PR 17-NOV-2000; 2000US-0249210.

PR 17-NOV-2000; 2000US-0249211.

PR 17-NOV-2000; 2000US-0249212.

PR 17-NOV-2000; 2000US-0249213.

PR 17-NOV-2000; 2000US-0249214.

PR 17-NOV-2000; 2000US-0249215.

PR 17-NOV-2000; 2000US-0249216.

PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI: 2001-488782/53.
XX
XX New polynucleotides and polypeptides for diagnosing, treating,
PT preventing or prognosing e.g. diseases or disorders of the nervous,
PT musculoskeletal, excretory, gastrointestinal, reproductive, and
PT respiratory systems
XX
XX
PS Disclosure: SEQ ID No 2132; 642pp; English.
XX
XX The invention relates to novel nucleic acids encoding novel human foetal
CC antigens. The nucleic acids and proteins are used to prevent, treat (e.g.
CC by gene therapy) or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC are also used in diagnosing a pathological condition or susceptibility
CC to a pathological condition. The antibodies to the antigens can also
CC be used in alleviating symptoms associated with the disorders and in
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC and ocular disorders e.g. corneal infection. The polypeptides can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. Numerous
CC examples of diseases and disorders treated by the nucleic acids and
CC proteins are given in the specification. The present sequence
CC is a genomic DNA fragment from a gene encoding a foetal antigen of the

Query Match 3.0%; Score 59.4; DB 22; Length 14809;
Best Local Similarity 81.2%; Pred. No. 8.1e-06;
Matches 69; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1917 AAGCCATTATACAAACACACATATATACACATATACACACACACACACATAT 1916
DB 10578 ATGGCATTATATATACACACACACACACACACACACACACACATATAT 10637
QY 1977 ATATATGAT 10637
DB 10638 AT 10662

RESULT 13
ABN81131/C
ID ABN81131 standard; DNA; 849 BP.
XX
XX
XX ABN81131;
AC
AC 16-JUL-2002 (first entry)
DT
DT Shrimp polynucleotide SEQ ID NO 135.
XX
XX
XX Giant black tiger prawn; Penaeus monodon; pacific white shrimp;
KM Litopenaeus vannamei; shrimp; microsatellite sequence; genome mapping;
KM Taura Syndrome Virus; TSV; infection; ds.
XX
XX
XX Litopenaeus vannamei.
OS
XX WO200034476-A2.
PN
XX
XX 15-JUN-2000.
PD
XX
XX 10-DEC-1999; 99WO-US29571.
PF
XX
XX 10-DEC-1998; 98US-0111670.
PR
XX
XX (TUFT) TUFTS COLLEGE.
PA
XX
XX Alciivar-Warren A, Xu Z, Dhar AK, Fan Y, Meehan D, Garcia DK;
PI
XX WPI: 2000-423422/36.
DR
XX
XX Polynucleotides of shrimp are useful for identifying, mapping and
PT characterizing of the genome of various species of shrimp
PT
XX
XX
XX Claim 1; Page 114; 120pp; English.
PS
XX
XX The invention relates to an isolated polynucleotide (I) of the giant
CC black tiger prawn, Penaeus monodon or expressed sequence tags of the
CC pacific white shrimp, Litopenaeus vannamei (ABN80997-ABN81172), both
CC containing microsatellite sequences including those P. monodon
CC microsatellite sequences given in GenBank AF077550-AF077598. (I), the
CC complementary sequence or fragment and the encoded polypeptide are useful
CC for mapping of the genome of various species of shrimp. Mapping the
CC genome of Penaeus is useful for determining whether a test shrimp,
CC preferably Litopenaeus vannamei, has a genotype associated with a
CC phenotypic trait such as resistance to Taura Syndrome Virus (TSV)
CC infection.
XX
XX Sequence 849 BP; 121 A; 113 C; 65 G; 152 T; 398 other;

Query Match 3.0%; Score 59.2; DB 21; Length 849;
Best Local Similarity 81.7%; Pred. No. 1.9e-06;
Matches 67; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1920 CCCAATTATACAAACACACATATATATACACATATACACACACACACATAT 1979
DB 253 CACANACANATATA 194
QY 1980 TATGAT 2001
DB 193 TAT 172

RESULT 14
ABL88366
ID ABL88366 standard; cDNA; 1721 BP.
XX
XX
XX ABL88366;
AC
XX
XX 16-MAY-2002 (first entry)
DT
XX
XX Pain regulated cDNA sequence 9.

XX Pain; analgesic; gene therapy; neurological disorder;
 KW neurodegenerative disease; gene; ss.
 XX Rattus sp..
 OS WO200212338-A2.
 PN 14-FEB-2002.
 PD 03-AUG-2001; 2001WO-EP09011.
 PE 03-AUG-2000; 2000DE-1037759.
 PR (CHEF) GRUENENTHAL GMBH.
 PA Gillen C, Wetzeis I, Whendt S, Weihe E, Schaefer MK.
 PI WPI: 2002-257469/30.
 DR Identifying pain-regulating compounds, useful for treating chronic pain
 XX and for diagnosis, by measuring binding of compounds to specific
 PT peptides and proteins -
 PS Claim 14; Fig 5; 213pp; German.
 XX The invention relates to identifying pain-regulating substances (A)
 CC comprises (i) incubating a test substance with a cell (or preparation
 CC from it) that has synthesised a peptide or protein (B) and (ii) measuring
 CC either binding of the test substance to (B) or some functional parameter
 CC that is altered by this binding. The method is useful for identifying
 CC pain-regulating substances (A) with analgesic activity. (A) along with
 CC nucleic acid (AB188411-AB188441) that encode proteins (B),
 CC AB885006-AB885037) that interact with (A); (B): vectors containing the
 CC nucleic acid; antibodies against (B); cells that express (B) and agents
 CC that bind to (B), are all useful for treating pain, particularly chronic
 CC pain, including use in gene therapy. The same materials can also be used
 CC for diagnosis, e.g. of neurological and neurodegenerative diseases. The
 CC present sequence is that of a polynucleotide of the invention.
 XX SO Sequence 1721 BP; 605 A; 342 C; 264 G; 457 T; 53 other;
 Query Match 3.0%; Score 59.2; DB 24; Length 1721;
 Best Local Similarity 54.0%; Pred. No. 2.8e-06;
 Matches 109; Conservative 5; Mismatches 88; Indels 0; Gaps 0;
 OY 1420 TACAGAGAGAGACTTACATGGACCTTGTAAAAAGCATAGGACATCACTAATAAGT 1479
 DB 895 TAGAAGTTTAAATCATATGAGAGACAAATTTCAACAAAGGTAGACAGAACTGAGAA 954
 OY 1480 TCAAGAAGTAAACATCATGTGTGAGTGAACAAGACATGGCCATGTTTTTTGTATG 1539
 DB 955 AAAACACCCAGATTTTCTGTACTTCTATATTGATACATGTTACTACTGTAAGAAA 1014
 OY 1540 AAACACAGCAGCAGGACAGGACACTGCTGCGACAGCGGCACACACACACAGCCGC 1599
 DB 1015 ATACACATGCAAGGCGWCAAGCGCGCGCGCGCGCGCGCGCACACACACACACAC 1074
 OY 1600 ACACACACACACACGATCAGC 1621
 DB 1075 ACACACACACACACACACACAC 1096

RESULT 15
 AAT8014/c
 ID AAT8014 standard; DNA; 6727 BP.

AC AAT8014;
 XX
 DT 22-APR-1998 (first entry)
 XX
 DE Murine IL-5 cDNA genomic fusion gene.
 XX

KW Murine CD3delta; transcription control sequence; transgenic mouse; TCS;
 KW T cell; interleukin-5; IL-5; thymocyte; disease model; therapeutic agent;
 KW screening; eosinophil; chimeric; fusion gene; ss.
 XX Chimeric - Mus sp.
 OS WO9738086-A2.
 PN 16-OCT-1997.
 PD 09-APR-1997; 97WO-US05932.
 PE 09-APR-1996; 96US-0629643.
 PR (MAYO-) MAYO FOUNDATION.
 PA Lee JT, Lee NA;
 PI WPI: 1997-512706/47.
 DR Transgenic mouse with interleukin-5 gene integrated into its genome
 XX - under control of thymocyte and T cell, lung or basal keratinocyte
 PT specific regulator, useful as disease model or to screen therapeutic
 PT agent
 PS Example 2; Page 73-75; 127pp; English.
 XX This is a murine interleukin-5 (IL-5) cDNA genomic fusion gene. This
 CC transgenic construct contain a chimeric pIL-5.KRPMNA sequence, comprising
 CC thymocyte/T cell, lung or basal keratinocyte specific transcription
 CC control sequence (TCS) and a segment encoding interleukin-5 (IL-5), which
 CC lacks endogenous 5'-control sequences when the TCS is thymocyte/T cell
 CC specific. The TCS is usually from a murine CD3delta gene. The chimeric
 CC DNA sequence is integrated into the genome and causes IL-5 to be
 CC expressed at a eosinophil related disease inducing level, specifically a
 CC serum level at least 50 times greater than the wild type. The mice can be
 CC used as models of IL-5 associated disease, e.g. asthma, eosinophilic
 CC pneumonia, myalgia, Loeffler's syndrome, allergy, emphysema, pulmonary
 CC fibrosis, Wegener's granulomatosis, adult respiratory distress syndrome
 CC (ARDS), bacterial or fungal infection, leukaemia, rheumatoid arthritis,
 CC atopic dermatitis or contact hypersensitivity, or to screen for IL-5
 CC antagonists which are potential therapeutic agents for an IL-5 associated
 CC disease. These mice constitutively express IL-5 at high levels in
 CC specific tissues.
 XX SO Sequence 6727 BP; 1765 A; 1538 C; 1414 G; 2010 T; 0 other;
 Query Match 3.0%; Score 59.2; DB 18; Length 6727;
 Best Local Similarity 83.8%; Pred. No. 6e-06;
 Matches 67; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
 OY 1922 CAATTATCAACACACACATATATACACATATACACACACACACACATATATATA 1981
 DB 2653 CATACATCATCATCATATATATATATACACACACACACACACACACATATATA 2594
 OY 1982 TGTATATATATATATACATA 2001
 DB 2593 TATATATATATATATATATA 2574

Search completed: February 17, 2003, 09:18:09
 Job time : 477.434 secs

AC ABL68253;
XX
DT 15-MAY-2002 (first entry)
XX
DE Kidney cancer related gene sequence SEQ ID NO:6590.
XX
XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
XX stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
XX cytostatic; gene therapy; antineoplastic; Wilms' tumour; adenocarcinoma;
XX gene; ds.
OS Homo sapiens.
XX
XX WO200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US10838.
XX
XX 05-JUN-2000; 2000US-209473P.
XX 05-JUN-2000; 2000US-209531P.
XX 18-SEP-2000; 2000US-233133P.
XX 18-SEP-2000; 2000US-233617P.
XX 20-SEP-2000; 2000US-234009P.
XX 20-SEP-2000; 2000US-234034P.
XX 22-SEP-2000; 2000US-234052P.
XX 22-SEP-2000; 2000US-234509P.
XX 22-SEP-2000; 2000US-234567P.
XX 25-SEP-2000; 2000US-234923P.
XX 25-SEP-2000; 2000US-234923P.
XX 25-SEP-2000; 2000US-235077P.
XX 25-SEP-2000; 2000US-235082P.
XX 25-SEP-2000; 2000US-235134P.
XX 26-SEP-2000; 2000US-235280P.
XX 26-SEP-2000; 2000US-235637P.
XX 27-SEP-2000; 2000US-235638P.
XX 27-SEP-2000; 2000US-235711P.
XX 27-SEP-2000; 2000US-235720P.
XX 27-SEP-2000; 2000US-235840P.
XX 27-SEP-2000; 2000US-235863P.
XX 28-SEP-2000; 2000US-236028P.
XX 28-SEP-2000; 2000US-236032P.
XX 28-SEP-2000; 2000US-236033P.
XX 28-SEP-2000; 2000US-236034P.
XX 28-SEP-2000; 2000US-236109P.
XX 28-SEP-2000; 2000US-236111P.
XX 29-SEP-2000; 2000US-236842P.
XX 29-SEP-2000; 2000US-236891P.
XX 02-OCT-2000; 2000US-237172P.
XX 02-OCT-2000; 2000US-237173P.
XX 02-OCT-2000; 2000US-237278P.
XX 02-OCT-2000; 2000US-237294P.
XX 02-OCT-2000; 2000US-237295P.
XX 02-OCT-2000; 2000US-237316P.
XX 03-OCT-2000; 2000US-237425P.
XX 03-OCT-2000; 2000US-237598P.
XX 03-OCT-2000; 2000US-237604P.
XX 03-OCT-2000; 2000US-237606P.
XX 03-OCT-2000; 2000US-237608P.
XX 01-NOV-2000; 2000US-244867P.
XX 01-NOV-2000; 2000US-245084P.
XX
XX (AVAL-) AVALON PHARM.
XX
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
XX Soppet DR, Weaver Z;
XX
XX MPI; 2002-188264/24.
XX
XX Screening for anti-neoplastic agent involves exposing cells to a
XX chemical agent to be tested for anti-neoplastic activity, and
XX determining a change in expression of a gene of a signature gene set
XX

PS Claim 1; SEQ ID 6590; 44pp; English.
XX
CC The present invention describes a method (M1) for screening for an
CC anti-neoplastic agent. The method involves exposing cells to a chemical
CC agent to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening
CC an anti-neoplastic agent, and can be used for producing a product which
CC is the data collected with respect to the anti-neoplastic agent as a
CC result of M1, and the data is sufficient to convey the chemical
CC structure and/or properties of the agent. M1 can be used in the
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Wilms' tumour.
XX
SQ Sequence 2353 BP; 493 A; 692 C; 673 G; 495 T; 0 other;

Query Match 21.1%; Score 283.8; DB 24; Length 2353;
Best Local Similarity 73.5%; Pred. No. 1.le-77;
Matches 392; Conservative 0; Mismatches 132; Indels 9; Gaps 2;

QY 816 GCAGAGCGGTGTTGATGATGACCAACACCGCCACTGACGGTGGATGCTGTCACA 875
DB 253 GCAGATGATGCTCTGATGATGATGACCAACATGACCTGACGGAGATGAGCCGTACG 312
QY 876 ACGTCTCTCCAGACACCGGCTTCACTGATGAGGCTGATGAGCATGATGAG 935
DB 313 ACGTGCACCTGTGACGAGAGGGCTTCAACCGCGCATGACCTGATGATGAG 312
QY 936 TGTGCTACCCATGATGACACACCTGCT--CCAAAGACGCTGTGATGACACCCGGCG 992
DB 373 TGCACCATTCCTGAGCTCACACATGCTCCGCAACAGCAGCTGCTAAACAGCCAGCG 432
QY 993 TCGTTAAGTGTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1052
DB 433 TCTTCT 492
QY 1053 GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1112
DB 493 GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 552
QY 1113 ACAGAGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1172
DB 553 GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 612
QY 1173 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1232
DB 613 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 666
QY 1233 AAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1292
DB 667 GCGCTCTGTGCGGAGATCCGTGTGACAGCGACCGACCTGAGACGATGAGCGACG 726
QY 1293 ACAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1345
DB 727 ACCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 779

RESULT 3
ABL68518
ID ABL68518 standard; DNA; 2353 BP.
XX
XX ABL68518;
XX
AC ABL68518;
XX
DT 15-MAY-2002 (first entry)
XX
DE Kidney cancer related gene sequence SEQ ID NO:6855.
XX

Query Match	21.1%	Score 283.8	DB 24	Length 2353
Best Local Similarity	73.5%	Pred. No. 1.1e-77		
Matches	392	Conservative	0	Mismatches 132; Indels 9; Gaps 2
QY	816	GCAGGACGGTGTGTAATGCCACAACCCACACCTTCACAGGTGATGTGTGTACA	875	
Db	253	GCAAGATGTGCTCTGATATGTACAGCAGCAATGCCACTGCACGGAGATGAGGCGTTACG	312	
QY	876	ACGTGCTCTCTCCACAGCCGGCTTCACTGATATGGGCTGTGTGTGAGACATGATATAG	935	
Db	313	ACGTGCACCTCTCTGAGAGGGCTTCAACCGGCATGGCCCTGGTGGACCTGGATATAG	372	
QY	936	TGTGTACCCCATGGAGTCAACACAGCTGCT---CCACACAGCTGTGTATACACCCCGGAC	992	
Db	373	TGCGCATCTCTGTGAGTCTACACATGCTCTCCGCAACAGACAGTGGTAAACAGCGCAAGC	432	
QY	993	TGCTTTAAGTCTCCTCTGTCAAGATGTGTTTGTGTGACGCGCTGAGCTGACAGTAT	1052	
Db	433	TCTCTCTCTCTGCTGTGTGCCCCGAAAGGCTTCCGCTGTGCGCGGCTGTGCTGTGCACAGAC	492	
QY	1053	GTGATATGATGCTCAGACAGAGGGGTCACTATCTATCTATGCCCCGTGGCCACTGTGTACAC	1112	
Db	493	GTGGATGTAGTCCGTCTGACACCTCGGGGTATGGACATCTGCCACGCCCTGTGGCCATGTGTCAAT	552	
QY	1113	ACAGAGAGCGACTACTTGTGTGTGTGTGTCCGACGAGGCTTTTACAGAGGGATGTGTACTGT	1172	
Db	553	GTGTGGGCGACTACTTGT	612	
QY	1173	GAGTGTCTCCCGACGAGCTCCTGTGTGAGCGACAGTGTGACTCTTTCGCCAGGCGCGATGGA	1232	
Db	613	GAGTGTCTCCCGGAGCTCCTGTGTGAGCGAGGAGTGTGACTCTTTCGCCAGGCGCGATGGA	666	
QY	1233	AAGCTGTGTGTCAAGACCCCTGCAATACATATGAGACCCCTGTAGATATCTGGCGCAGC	1292	
Db	667	GCGCTCTGT	726	
QY	1293	ACAGAGT	1345	
Db	727	ACCGAGT	779	

RESULT 4

ABL68872

ID ABL68872 standard; DNA; 2353 BP.

XX ABL68872;

XX 15-MAY-2002 (first entry)

DE kidney cancer related gene sequence SEQ ID NO:7209.

XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;

KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;

KM cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;

gene; ds.

XX Homo sapiens

DS

XX WO200194629-A2.
 PN
 XX
 PD 13-DEC-2001.
 XX
 PF 30-MAY-2001; 2001MO-US106838.
 XX
 PR 05-JUN-2000; 2000US-209473P.
 PR 05-JUN-2000; 2000US-209531P.
 PR 18-SEP-2000; 2000US-233133P.
 PR 18-SEP-2000; 2000US-233617P.
 PR 20-SEP-2000; 2000US-234009P.
 PR 20-SEP-2000; 2000US-234034P.
 PR 20-SEP-2000; 2000US-234052P.
 PR 22-SEP-2000; 2000US-234509P.
 PR 22-SEP-2000; 2000US-234567P.
 PR 25-SEP-2000; 2000US-234923P.
 PR 25-SEP-2000; 2000US-234924P.
 PR 25-SEP-2000; 2000US-235077P.
 PR 25-SEP-2000; 2000US-235082P.
 PR 25-SEP-2000; 2000US-235134P.
 PR 25-SEP-2000; 2000US-235280P.
 PR 26-SEP-2000; 2000US-235638P.
 PR 27-SEP-2000; 2000US-235711P.
 PR 27-SEP-2000; 2000US-235720P.
 PR 27-SEP-2000; 2000US-235840P.
 PR 28-SEP-2000; 2000US-236028P.
 PR 28-SEP-2000; 2000US-236032P.
 PR 28-SEP-2000; 2000US-236033P.
 PR 28-SEP-2000; 2000US-236034P.
 PR 28-SEP-2000; 2000US-236109P.
 PR 29-SEP-2000; 2000US-236111P.
 PR 29-SEP-2000; 2000US-236842P.
 PR 29-SEP-2000; 2000US-236891P.
 PR 02-OCT-2000; 2000US-237172P.
 PR 02-OCT-2000; 2000US-237173P.
 PR 02-OCT-2000; 2000US-237278P.
 PR 02-OCT-2000; 2000US-237294P.
 PR 02-OCT-2000; 2000US-237295P.
 PR 03-OCT-2000; 2000US-237316P.
 PR 03-OCT-2000; 2000US-237425P.
 PR 03-OCT-2000; 2000US-237598P.
 PR 03-OCT-2000; 2000US-237604P.
 PR 03-OCT-2000; 2000US-237606P.
 PR 03-OCT-2000; 2000US-237608P.
 PR 01-NOV-2000; 2000US-244867P.
 PR 01-NOV-2000; 2000US-245084P.
 XX
 PA (AVALON PHARM.
 XX
 PI Young PE, Augustus M, Carter KC, Edner R, Endress G, Horrigan S,
 PI Soppet DR, Weaver Z;
 XX
 DR WPI; 2002-188264/24.
 XX
 PT Screening for anti-neoplastic agent involves exposing cells to a
 PT chemical agent to be tested for anti-neoplastic activity, and
 PT determining a change in expression of a gene of a signature gene set -
 XX
 PS Claim 1; SEQ ID 7209; 44pp; English.
 XX
 CC The present invention describes a method (M1) for screening for an
 CC anti-neoplastic agent. The method involves exposing cells to a chemical
 CC agent to be tested for anti-neoplastic activity, determining a change in
 CC expression of at least one gene (I) of a signature gene set, where (I)
 CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
 CC to ABL70110), or is at least 95% identical to (S), where a change in
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
 CC activity and can be used in gene therapy. M1 can be used for screening
 CC an anti-neoplastic agent, and can be used for producing a product which
 CC is the data collected with respect to the anti-neoplastic agent as a

CC result of M1, and the data is sufficient to convey the chemical
 CC structure and/or properties of the agent. M1 can be used in the
 CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
 CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
 CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
 CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
 CC carcinoma, papillary carcinoma and Wilms' tumour.
 XX
 SQ Sequence 2353 BP; 493 A; 692 C; 673 G; 495 T; 0 other:
 Query Match 21.1%; Score 283.8; DB 24; Length 2353;
 Best Local Similarity 73.5%; Pred. No. 1.1e-77;
 Matches 392; Conservative 0; Mismatches 132; Indels 9; Gaps 2;
 QY 816 GCAGAGCGGTGTTCTGAATGCCACACACGACCTGACAGGTGATGTTGTCACA 875
 DB 253 GCAGATGGTGTCTGATGATGCACAGCAATGCACCTGCAGAGATGAGCCGTTACG 312
 QY 876 ACGTGCTCCTGGCAGACCGGCTTCACGTGATGAGGGGTGTGTGTGATGATGATGAG 935
 DB 313 ACGTGCACTGTTCAGAGAGGCTTCACCGGATGCTGACCTGAGACTGTGATGAG 372
 QY 936 TGTGCTACCCCATGAGACTCAACAATGCT--CCACAGCAGCTGTGTAACCCCGGCG 992
 DB 373 TCGGCATTCCTGAGACTCAACAATGCTCCGCAACGACGACTGCTGATGAG 432
 QY 993 TCGTTAACTGCTCCTGTCAGATGTTTTCGTCAGCGCTGAGCTGACCTGAT 1052
 DB 433 TCGTCTCCTGCTGCTGCCCCGAGAGGCTTCGCTGCGCCGGTCTCGGCTGCACAGAC 492
 QY 1053 GTGATGATGCTGCAGACAGAGGGGCTCAGTAATGTCATGCGCGCCACCTGTGTACAC 1112
 DB 493 GTGATGATGCTGCTGAGAGGCTTCAGGCTTACCTGACCTGCCACCCCTGTCATAT 552
 QY 1113 ACAGAGGAGGACTACTGTGCGTGTGTCGAGGAGGCTTTACAGGGATGTTGTAAGT 1172
 DB 553 GTGATGAGGAGGACTACTGTGCGTGTGTCGAGGAGGCTTTACAGGGAGTGTGTAAGT 612
 QY 1173 GAGTCTCCGCCAGGCTCCTGTCAGACGAGACTGACTGCTGCCACGAGGCGGATGGA 1232
 DB 613 GAGTCTCCGCCAGGCTCCTGTCAGGCGGAGGTTGACTGTGCTGCCGAGGAGC-----GAC 666
 QY 1233 AAGCTGATGTCAGAGACCCCGCAATATGATGATGATGATGATGATGATGATGATG 1292
 DB 667 GCGCTCGTGTGCGGATTCGTCAGGCGACCCGACCCCTGACAGATGATGATGATGATG 726
 QY 1293 ACAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1345
 DB 727 ACCGATACGGGAGGAGGCTACCGCTGCGACAGGAGACTGCGGCTGTGTACCG 779
 RESULT 5
 AAH57516
 ID AAH57516 standard; cDNA; 2439 BP.
 XX
 AC AAH57516;
 XX
 DT 10-SEP-2001 (first entry)
 XX
 DE Human kidney cell specific cDNA sequence SEQ ID NO:356.
 XX
 KW Human; tissue specific; diagnosis; brain; heart; skeletal muscle;
 KW lung; liver; uterus; ovary; stomach; intestine; kidney; pancreas; ss;
 KW metabolic disease; developmental disease; cytostatic; immunomodulatory;
 KW neuroprotective; gene therapy; cancer; immunopathology; neuropathology.
 OS Homo sapiens.
 XX
 PN WO200132927-A2.
 XX
 PD 10-MAY-2001.
 XX
 PF 02-NOV-2000; 2000MO-US30396.

XX 04-NOV-1999; 99US-0163508.
 XX
 XX (INCY-) INCYTE GENOMICS INC.
 XX
 XX Sornasse T, Sellhauer JJ, Watson GA;
 XX WPI, 2001-291057/30.
 XX
 XX New cell and tissue specific polynucleotides useful for diagnosis,
 XX prognosis or monitoring of treatments for disorders where the gene is
 XX associated with a cancer, immunopathology or neuropathology -
 XX
 XX Claim 1; Page 275-276; 327pp; English.
 XX
 XX AAH57161 to AAH57576 represent cell and tissue specific polynucleotide
 XX sequences (I). (I) can have cytostatic, immunomodulatory and
 XX neuroprotective activities, and can be used in gene therapy. (I) and
 XX proteins (II) encoded by then are used in high throughput screening
 XX assays to select DNA molecules, RNA molecules, peptide nucleic acids,
 XX mimetics, peptides, proteins, agonists, antagonists, antibodies or
 XX their fragments, immunoglobulins, inhibitors, drug compounds and
 XX pharmaceutical agents. Expression of (I) in a sample indicates the
 XX differentiation of embryonic stem cells into a tissue selected from
 XX brain, heart, kidney, liver, lung, skeletal muscle or pancreatic
 XX tissues. (I) and (II) are used to produce an expression profile that
 XX defines a metabolic or developmental process, treatment, condition,
 XX disease or disorder. The gene profile can be used for diagnosis,
 XX prognosis or monitoring of treatments and for investigating a
 XX predisposition to a disorder where the gene is associated with a
 XX cancer, immunopathology or neuropathology.
 XX

SO Sequence 2439 BP; 524 A; 707 C; 694 G; 513 T; 1 other;

Query Match 21.1%; Score 283.8; DB 22; Length 2439;
 Best Local Similarity 73.5%; Pred. No. 1.1e-77;
 Matches 392; Conservative 0; Mismatches 132; Indels 9; Gaps 2;

QY 816 GCAGAGCGGTGTTCTGAAATGCCAACACGCGCACTGCGAGGTGGTGTGTCACA 875
 DB 329 GCMAATGTTGCTGCTGAAATGCTACACACGCAACGCGAGGTGAGCGCGTACG 388
 QY 876 ACGTCTCTGCGAGCGCGCTGCTGAGTGGGCTGCTGAGTGGAGTGGATGAG 935
 DB 369 ACGTGCACCTGTGAGGAGGCTTACCGCGGATGGCTGACCTGCGTGGATGATGAG 448
 QY 936 TGTGCTACCCCATGAGACTACAACTGCT---CCAAACAGACGCTGTGTAACCCCGGCG 992
 DB 449 TCGGCAATTCCTGAGAGCTACAACTGCTCGGCAACGAGAGCTGTAACAGCGCAGGC 508
 QY 993 TCGTTTAACTGCTCTGTCAGAGTGGTTTCTGTCAGCGCTGAGCTGACATGAT 1052
 DB 509 TCCCTTCTCTGCTGCTGCCCGGAAAGGCTTCCGCTGTGCGCCGCTGCTGACAGAC 568
 QY 1053 GGGATGAGTGCACAGAGAGGCGGCTGATGATGATGATGATGATGATGATGATGAT 1112
 DB 569 GGGATGAGTGCAGAGAGGCTGATGATGATGATGATGATGATGATGATGATGATGAT 628
 QY 1113 ACAGAGGCGACTACTTGTGCGTGTGCCGAGGCTTTTACAGGAGTGTGTTGATCTGT 1172
 DB 629 GGGTGGGCGAGCTACTTGTGCGTGTGCCGAGGCTTTTACAGGAGTGTGTTGATCTGT 688
 QY 1173 GAGTGTGCGCGAGGCTGCTGTGAGCAGAGCTGATGCTTGGCCCAAGGCGCGAGGGA 1232
 DB 689 GAGTGTGCGCGAGGCTGCTGTGAGCAGAGCTGATGCTTGGCCCAAGGCGCGAGGGA 1292
 QY 1233 AAGTGTGCTGTCAGAGCCCTGATGATGATGATGATGATGATGATGATGATGATGAT 1292
 DB 743 GCGTGTGCTGTCAGAGCCCTGATGATGATGATGATGATGATGATGATGATGATGAT 802
 QY 1293 ACAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1345
 DB 803 ACCGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 855

RESULT 6
 AA293969
 ID AA293969 standard; cDNA; 655 BP.

AA293969;

29-AUG-2000 (first entry)

Goat uromodulin gene exon 1.

Uromodulin; promoter; kidney; urine; heterologous gene; treatment;
 therapy; gene expression; pharmaceutical; goat; ss.

Capra hircus.

WO200029608-A1.

25-MAY-2000.

12-NOV-1999; 99WO-US26870.

13-NOV-1998; 98US-0108195.

09-JUL-1999; 99US-0142925.

(UYNY) UNIV NEW YORK STATE.

Wu X, Sun T;

WPI; 2000-387816/33.

New kidney-specific promoter useful for production of transgenic
 animals as urinary bioreactors, is operably linked to a heterologous
 gene

Example 2; Fig 10b; 55pp; English.

New methods to produce heterologous recombinant proteins in urine
 require the use of a DNA molecule which is a kidney-specific
 promoter, such as the uromodulin promoter, operably linked to a
 heterologous gene encoding a biologically active protein. The
 uromodulin promoter expresses the heterologous gene in vivo in the
 kidneys to produce a recombinant biologically active protein in the
 urine. The recombinant proteins produced may be useful for treating
 human diseases. The major advantages of using this urine-based system
 over milk-based systems are the ability to harvest the product soon
 after birth and throughout the life of the animal irrespective of sex
 or reproductive status, and the ease of product purification from
 urine. In addition, livestock urine is a proven, currently utilized
 source of pharmaceuticals.

Sequence 655 BP; 116 A; 187 C; 211 G; 138 T; 3 other;

Query Match 18.9%; Score 254.2; DB 21; Length 655;
 Best Local Similarity 70.3%; Pred. No. 9.3e-69;
 Matches 400; Conservative 0; Mismatches 158; Indels 11; Gaps 4;

QY 779 CCAGCTGACGAGGATCTGAGACTGGGCTCTTCTGCCGACGAGGAGCTTCTGATGCCA 838
 DB 94 CAAGGGGTCCTCCACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 153
 QY 839 CAACCAAGCCACTGACAGGTGATGTC-TGCTACACAGTCTCTCTCCAGACCGGCT 897
 DB 154 CCGCAATGACACTTGTACGCTGACGAGGCGGCTTGCACAGACTGCGCTGCAGAGGCGCT 213
 QY 898 TCACGTGATGAGGCTGGTGTGAGAGCAATGATGATGATGATGATGATGATGATGATGAT 957
 DB 214 TCACT-GCGACGCGCTCGAATGTCGATGATGATGATGATGATGATGATGATGATGATGAT 272
 QY 958 ACTGCT---CCAACAGACGCTGTGTAACACCCCGGCGCTGTTAAGTGTCTCTGTGAGG 1014
 DB 273 ACTGCTCCGCCACCAACAGCTGTGTAACCCCGGCGCTGTTAAGTGTCTCTGTGAGG 332

QY	1015	ATGCTTTTCGTCCTGACGCGCTTGACGCTGCACGTGATGTGATGATGCTGCAGAGCAGG	1074
Db	333	AAAGTTTTCCTCTGAGCTTCGGAGCTCGGCTCGAGAGATGTGGACAGGTGTGCAGAGCCAG	392
QY	1075	GGCGCAGTAATCTGTCAATGTCGCCCTGGCCGACCTGTGTCAACACAGAGAGCGACTACTTGTGCG	1134
Db	393	GGCTCAGCCCGGTGCGACGCCCTCGCGCACCCTGTGATCAATGGCGAGGCGCACTACTCAATGCG	452
QY	1135	TGTGTCCCGAAGGGGCTTTCACAGGGATGTGTGTACTGATGATGCTCCCGAGGCTCCGTG	1194
Db	453	TGTGTCCCGGCGGGCTGACGTGGGGAGGAGGAGCAGTGTAGTGTTCCTCCGCGCTCTGCG	512
QY	1195	AGCCAGAGACTGGAATGCTTGTGCCGCCACGGGCCCGGATGAAAGCTGTGTGTCAAGACCCCT	1254
Db	513	GGCGCTGGGCTAAGACTGGGTGCGGGAGGGG-----TGACCGCGTAGTGTGGCGGTGACCCGT	566
QY	1255	GCATATACATTTGAGACCCCTGACTGATGATGCGGCGAGCAGCAGATATGTTGTGGCTACT	1314
Db	567	GCCAGGGCGCACACATCTGTGACGATATCTGGCGCAGCAGAGATACGGCTCCGGCTACG	626
QY	1315	CCGTGTGACGCGGGCTGTGACGCGCTGGTAC	1343
Db	627	TCTGTGATGTCAAGTCTGGGCGGTGTGAC	655
RESULT 7			
ACG04147			
ID	AAC04147 standard; cDNA; 408 BP.		
AC	AAC04147;		
XX			
DT	06-OCT-2000 (first entry)		
XX			
DE	Human secreted protein 5' EST, SEQ ID NO: 8222.		
XX			
KW	Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation,		
KM	gene therapy; chromosome mapping; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	EP1033401-A2.		
XX			
PD	06-SEP-2000.		
XX			
PE	21-FEB-2000; 2000EP-0200610.		
XX			
XX	26-FEB-1999; 1990S-0122487.		
PR	(GEST) GENSET.		
XX			
PA	Dumas Milne Edwards J, Duclert A, Giordano J;		
XX			
PI	WPI; 2000-500381/45.		
XX			
DR			
XX			
PT	New nucleic acid that is a 5' expressed sequence tag (5' EST) for		
PT	obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for		
PT	diagnostic, forensic, gene therapy and chromosome mapping procedures -		
XX			
XX	Claim 1: SEQ ID 8222; 71bp + CD-ROM; English.		
CC			
CC	The present sequence is one of a large number of 5' ESTs derived from		
CC	mRNAs encoding secreted proteins. No ORF has yet been conclusively		
CC	identified within the present sequence. The 5' ESTs were prepared from		
CC	total human RNAs or polyA+ RNAs derived from 30 different tissues. EST		
CC	sequences usually correspond mainly to the 3' untranslated region (UTR)		
CC	of the mRNA because they are often obtained from oligo-dT primed cDNA		
CC	libraries. Such ESTs are not well suited for isolating cDNA sequences		
CC	derived from the 5' ends of mRNAs and even in those cases where longer		
CC	cDNA sequences have been obtained, the full 5' UTR is rarely included.		
CC	5' ESTs are derived from mRNAs with intact 5' ends and can therefore be		
CC	used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used		
CC	in diagnostic, forensic, gene therapy and chromosome mapping procedures.		

CC	They are used to obtain upstream regulatory sequences and to design expression and secretion vectors.
CC	
XX	
SQ	Sequence 408 BP; 96 A; 112 C; 113 G; 83 T; 4 other;
	Query Match 6.9%; Score 93.4; DB 21; Length 408; Best Local Similarity 73.6%; Pred. No. 1.4e-18; Matches 145; Conservative 0; Mismatches 48; Indels 4; Gaps 2;
OY	816 GCAGGACGGTGTTCATGCACAACACCGCACCCTGCACGGTGATGGTTGTCACA 875
Db	213 GCAGAATGGTGCTCTGTAATGTCACAGCAATGCCACTTCGACGGAGATGAGGCCATTACG 272
OY	876 ACGTGCTCCTCCACAGACCGGCTTCACCTGGTATGGGCTGGTGTGAGACATGATGAG 935
Db	273 ACGTGACCTGTGTAGSAGGGCTTCACCGGCGCATGGCTCACTGGTGACCTGGATGAG 332
OY	936 TTGTGCTAACCCCATGGACTCACACTGCT---CCAACAGCAGTGTGTGAACACCCGGGC 992
Db	333 TTGGGC-ATTCTGTGAGGCTCACAACTGCTCCGCCAACAGCAGCTGCTAAACAGCCAGNT 391
OY	993 TCGTTTAAGTGCTCCTG 1009
Db	392 TCCTTCTCTGCNCCTG 408
RESULT 8	
AAO03324	
ID	AAO03324 standard; DNA: 6567 BP.
XX	AAO03324;
AC	
XX	
DT	02-AUG-1990 (first entry)
XX	
DE	Eimeria tenella genomic DNA encoding antigen GX5401FL.
KW	Eimeria tenella; antigen GX5401FL; antigen GX5401; avian coccidiosis.
XX	
OS	Eimeria tenella.
FH	Key Location/Qualifiers
FT	CDS 1..6567
FT	/tag= a
PN	WO9000403-A.
XX	
PD	25-JAN-1990.
XX	
PF	05-JUL-1989; 89WO-US02918.
XX	
PR	05-JUL-1988; 88US-0215162.
XX	
PA	(GENE-) GENEX CORP.
XX	
PI	Anderson DM, McCandliss RJ, Strausberg SL, Strausberg RL;
DR	MP1: 1990-051586/07.
DR	P-PSDB: AAR05222.
XX	
PT	Cloned gene or fragment encoding antigenic protein -
PT	which binds with antibodies against avian coccidia, and
PT	transformed cells used in vaccine
PS	Claim 10; Page 93 and Fig 14; 134pp; English.
CC	
XX	An E. tenella genomic library was screened with radioactively labeled cDNA encoding the GX5401 antigen to identify clones encoding extensions of the GX5401 coding sequence. Several plaques reactive with the cDNA were identified. DNA from clone number 533 was sequenced. This analysis showed that the coding sequence in this clone overlaps with the GX5401 coding sequence and extends that sequence toward the N-terminal coding region. Analysis of the new sequence together with the GX5401 antigen coding sequence reveals an open reading frame encoding an Rimeria

CC protein of about 250,000 daltons. The protein carries several repeated
CC peptide sequences and is rich in cysteine residues. The open reading
CC frame encodes a potential signal sequence for protein secretion. The
CC coding sequence for this antigenic protein is given in AAN93334. Also new
CC are an expression vector conty. the cloned gene, host cells transformed
CC with the vector, and AP encoded by the cloned gene. The transformed cells
CC are used in a vaccine to immunise birds against avian coccidiosis. By
CC labelling the peptides, they can be used as a type-specific probe. The
CC AP may also be used in an assay to detect Ab against the coccidia. The
CC Abs are used to identify transformed cells conty. the DNA.
XX
SQ Sequence 6567 BP; 1490 A; 1741 C; 2189 G; 1147 T; 0 other;
Query Match 4.7%; Score 63.8; DB 11; Length 6567;
Best Local Similarity 52.9%; Pred. No. 1.2e-08;
Matches 137; Conservative 0; Mismatches 122; Indels 0; Gaps 0;
OY 812 GTCCGAGGAGCGTGTTCGAATGCCAACCAACGACGCTGCACGCTGATGGTGTGTGT 871
DB 1173 GTGTGACGGGACGCGGCGAATGCCAGCTCCGACAGTGTGTGAACGTGGATGGCAG 1232
OY 872 CACAACGTGCTCTGCTGCACAGCGCTTCACTGTGATGGCTGTGTGAGACATAGA 931
DB 1233 CTATGAATGCCACTGTGTGGAAGGTTTCATGTGGCGAAGAGTGTGACAGTGA 1292
OY 932 TGATGTGTGCTACCCCATGAGTCAACACTGCTCCACAGCAGCTGTGTGAACCCCGGG 991
DB 1293 CGAGTGTGGGGCTGAGAGCTTCCGCTGTGGCCCAACACCATGCTCTGAACACATCG 1352
OY 992 CTCGTTAGTCTCTCTCTGACAGATGTTTCTGTGACGCTGAGCTGAGCTCACTGA 1051
DB 1353 CAGCTACGAGTGTGAGTGAAGGAGATATGCGCACATGAGGAGCAACGCGTCAAGCA 1412
OY 1052 TGTGATGAGTGTCTCAGAG 1070
DB 1413 CATGATGAGTGTCTCAGAG 1431
RESULT 9
AAV44292
ID AAV44292 standard; CDNA; 761 BP.
XX
AC AAV44292;
XX
DT 06-OCT-1998 (first entry)
XX
DE Human secreted protein clone CB107_1 CDNA.
XX
KW Secreted protein; human fetal brain; nutrition; cytokine; stimulant;
KW cell proliferation; differentiation; immune system; suppressor; ligand;
KW regulator; hematopoiesis; tissue growth; activin; inhibin; haemostatic;
KW chemotaxis; chemokinesis; thrombosis; receptor; cadherin; tumour;
KW anti-inflammatory; ds.
XX
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT 41..760
FT CDS /tag= a
FT /product= secreted protein
FT /note= "isolated from clone CB107_1"
XX
XX MO9827205-A2.
XX
XX 25-JUN-1998.
XX
XX 17-DEC-1997; 97WO-US23330.
XX
XX 16-DEC-1997; 97US-0991872.
XX 18-DEC-1996; 96US-0769192.
XX 13-JAN-1997; 97US-0783401.
XX
XX (GENY) GENETICS INST INC.

XX
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI Race LA, Spaulding V, Treacy M;
XX
XX WPI; 1998-362774/31.
DR P-PSDB; AAM64219.
XX
XX New polynucleotides and secreted proteins - obtained from human
PT foetal brain, human adult testes, human adult brain and human adult
PT salivary gland cDNA libraries
XX
PS Claim 1a; Page 64-65; 110pp; English.
XX
CC This sequence encodes a novel secreted protein from clone CB107_1
CC isolated from a human fetal brain cDNA library. This protein has
CC applications for nutritional use, cytokine and cell
CC proliferation/differentiation activity, immune stimulating or
CC suppressing activity, hematopoiesis regulating activity, tissue growth
CC activity, activin/inhibin activity, chemotactic/chemokinetic activity,
CC haemostatic and thrombotic activity, receptor/ligand activity,
CC anti-inflammatory activity, cadherin/tumour invasion suppressor activity,
CC tumour inhibition activity and other activities.
XX
SQ Sequence 761 BP; 205 A; 130 C; 200 G; 222 T; 4 other;
Query Match 4.4%; Score 59; DB 19; Length 761;
Best Local Similarity 55.1%; Pred. No. 1.1e-07;
Matches 135; Conservative 0; Mismatches 107; Indels 3; Gaps 1;
OY 918 TGTGAGGACATGAGTGTGCTACCCCATGAGTCAACACGCTCCACAGCAGCTGT 977
DB 257 TGTGAAGACATGATGAGTGTGCTGAAGGCGCATTAATGTTGAATACATATGTGT 316
OY 978 GTGAACACCCCGGCTGCTTTAAGTCTCTGTCAGAGATGTTTCTGTGACCCCTGAG 1037
DB 317 GTCAACACCCCGGCTCTTTATGTGATGTCATGTGCAAACTGATACATGATGAT 376
OY 1038 CTGAGCTGACATGATGTGATGATGCTCAGACAGAGGCTCAGTACTGATGCCCTG 1097
DB 377 TATTCATGTACAGACATGATGATGTATACCAATTCAGACAG--CTGTGATGAANAAT 433
OY 1098 GCCACCTGTGTACACAGAGGCGACTACTTGTGCTGTGCTCCGAGGCTTTACAGGG 1157
DB 434 GCTTTATGCTTCAACACTGTTGGAGACACACTGTGTTGCAAGCGGCTATACAGGG 493
OY 1158 GATGG 1162
DB 494 AATGG 498
RESULT 10
AAF98466
ID AAF98466 standard; CDNA; 761 BP.
XX
AC AAF98466;
XX
DT 07-JUN-2001 (first entry)
XX
DE Human cDNA clone CB107_1 internal sequence SEQ ID 155.
XX
KW Human; secreted protein; nutrient; cytokine modulator; proliferation;
KW differentiation; immune system modulator; tissue growth; chemotactic;
KW haemostatic; thrombolytic; anti-inflammatory; tumour inhibition; ss;
KW haematopoiesis.
XX
XX Homo sapiens.
XX
XX WO200119988-A1.
XX
XX 22-MAR-2001.
XX
XX 14-SEP-2000; 2000WO-US25135.
XX

PR 17-SEP-1999; 99US-0398829.
XX
XX (GEM) GENETICS INST INC.
PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racle LA, Evans C;
PI Merberg D, Treacy M, Bowman MR, Spaulding V, Agostino MJ;
DR MPI; 2001-244801/25.
DR P-PSDB; AAB90729.
XX
XX Isolated nucleic acids encoding polypeptides, useful for modulating
PT e.g. cytokine and cell proliferation/differentiation activity, the
PT immune system and hematopoiesis regulating activity -
XX
XX Disclosure; Page 483; 557pp; English.
PS
XX Human cDNA clones represented in AAF98374 - AAF98489 encode secreted
CC proteins AAB90667 - AAB90750. The cDNA clones are isolated from various
CC tissue types, and may be used in the prevention, treatment and diagnosis
CC of diseases associated with inappropriate protein expression. The
CC polypeptides and nucleic acids may be used as nutrients or to modulate
CC cytokine and cell proliferation/differentiation activity and may also be
CC involved in modulation of the immune system. The cDNA sequences,
CC proteins, their agonists and/or antagonists exhibit haematopoiesis
CC regulating activity; tissue growth activity; activin/inhibin activity;
CC chemotactic/chemokine activity; haemostatic and thrombolytic
CC activity; receptor/ligand activity; anti-inflammatory activity;
CC haematopoiesis activity; cadherin/tumour suppressor activity; and/or
CC tumour inhibition activity. Included in the invention are probes
CC represented in AAF98490 - AAF98572 which are specific for the cDNA clones
XX encoding the secreted proteins.
SQ
XX Sequence 761 BP; 205 A; 130 C; 200 G; 222 T; 4 other:
Query Match 4.4%; Score 59; DB 22; Length 761;
Best Local Similarity 55.1%; Pred. No. 1.1e-07;
Matches 135; Conservative 0; Mismatches 107; Indels 3; Gaps 1;
OY 918 TGTGAGACATGATGATGCTGCTACCCCATGACACAGCTGCTCCACAGCAGCTGT 977
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
257 TGTGAGACATGATGATGCTGCTGAGGCGCATCTACTGTGAAATACATGATGCT 316
OY 978 GTGAACACCCCGGGCTCGTTTAAGTCTCTCTGTCAGATGCTTTCCTGACCCCTGAG 1037
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
317 GTCAACACCCCGGGCTCTTTATGTCATCTGCAAACTGGATACATCAGATTGATGAT 376
OY 1038 CTGACCTGCATGATGATGATGATGCTGACAGCAGGGCTGCTGTAAGTCTGATGCCCTG 1097
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
377 TATTCATGTACAGACATGATGATGATGATGATGATGATGATGATGATGATGAT 433
OY 1098 GCCACCTGTGTCACACAGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1157
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
434 GCTTTATGCTTCACACAGCTGTTGAGAGACACAACTGTGTTGCAAGCGGGCTATACAGGG 493
OY 1158 GATGG 1162
DB |||||
494 AATGG 498
XX
XX RESULT 11
AAV01882
ID AAV01882 standard; cDNA; 2448 BP.
XX
XX AAV01882;
AC
XX 20-APR-1998 (first entry)
XX
XX Human nel-related protein type 2 gene.
XX
XX Human; foetal brain cDNA library; GDP dissociation stimulating protein;
KW brain specific nucleosome assembly protein; diagnosis; therapy;
KW skeletal muscle specific ubiquitin conjugating enzyme; TMF-2; NPIK;
KM nel-related protein type 1; nel-related type 2; hereditary disease;
XX

KW cancer; ss.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH 1..2448
FT CDS /lag= a
FT /product= "nel-related protein type 2"
FT /note= "no stop codon given"
XX
XX EP796913-A2.
XX
XX 24-SEP-1997.
XX
XX 19-MAR-1997; 97EP-0104842.
XX
XX 05-MAR-1997; 97JP-0069163.
XX 19-MAR-1996; 96JP-0063410.
XX
XX (SARA) OTSUKA PHARM CO LTD.
XX
XX Fujiwara T, Horie M, Watanabe T;
DR MPI; 1997-459830/43.
XX P-PSDB; AAM37501.
XX
XX Novel human genes, e.g. brain-specific nucleosome assembly protein -
PT useful for diagnosis or therapy of hereditary disease and cancer
XX
XX Claim 23; Page 107-109; 123pp; English.
XX
XX The present sequence encodes a nel-related protein type 2 isolated from
CC a human foetal brain cDNA library. The nucleotide or amino acid
CC sequences are useful for in-vitro diagnosis of hereditary diseases and
CC cancer and for preparation of pharmaceuticals.
XX
SQ
XX Sequence 2448 BP; 660 A; 521 C; 611 G; 656 T; 0 other:
Query Match 4.3%; Score 57.8; DB 18; Length 2448;
Best Local Similarity 55.1%; Pred. No. 5.3e-07;
Matches 135; Conservative 0; Mismatches 107; Indels 3; Gaps 1;
OY 918 TGTGAGACATGATGATGCTGCTACCCCATGACACAGCTGCTCCACAGCAGCTGT 977
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1312 TGTGAGACATGATGATGCTGCTGAGGCGCATCTACTGTGTAATACATGATGCT 1371
OY 978 GTGAACACCCCGGGCTCGTTTAAGTCTCTCTGTCAGATGCTTTCCTGACCCCTGAG 1037
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1372 GTCAACACCCCGGGCTCTTTATGTCATCTGCAAACTGGATACATCAGATTGATGAT 1431
OY 1038 CTGACCTGCATGATGATGATGATGCTGACAGCAGGGCTGCTGTAAGTCTGATGCCCTG 1097
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1432 TATTCATGTACAGACATGATGATGATGATGATGATGATGATGATGATGATGAT 1488
OY 1098 GCCACCTGTGTCACAGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1157
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1489 GCTTTATGCTTCACACAGCTGTTGAGAGACACAACTGTGTTGCAAGCGGGCTATACAGGG 1548
OY 1158 GATGG 1162
DB |||||
1549 AATGG 1553
XX
XX RESULT 12
AAV01883
ID AAV01883 standard; DNA; 3198 BP.
XX
XX AAV01883;
AC
XX 20-APR-1998 (first entry)
XX
XX Human nel-related protein type 2 gene.
XX


```

XX MOZ00168912-A2.
PN
PD 20-SEP-2001.
PF
PP 15-MAR-2001; 2001WO-EPO2955.
PR 15-MAR-2000; 2000DE-1013847.
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032528.
PR 01-SEP-2000; 2000DE-1043826.
PA (EPIC-) EPIDENOMICS AG.
PI Olek A., Piepenbrock C., Berlin K.;
PX WPl: 2001-602752/68.
PT Fragments of chemically modified genes associated with tumour suppressor
PT genes and oncogenes, useful in designing primers and probes for
PT analysing diseases associated with cytosine methylation state e.g.
PS cancer
PS Claim 1; SEQ ID No 55; 27pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulfite, of genes associated with tumour suppression and
CC oncogenes having a sequence taken from 536 (actually 533 since
CC numbers 408, 458 and 500 are missing from the sequence listing), sequences
CC (Ss) and sequences complementary to (Sp). The nucleic acid may be a
CC peptide nucleic acid-oligonucleotide (PNA). Of at least 9 nucleotides and may
CC form part of a set of probes for detecting the cytosine methylation state
CC and/or single nucleotide polymorphisms and also to be used in an
CC array for analysing diseases associated with CpG dinucleotides e.g.
CC cancers and tumours. The probes can also be used in a method for
CC ascertaining genetic and/or epigenetic parameters for the diagnosis
CC and/or therapy of existing diseases or the predisposition to specific
CC diseases, by analysing cytosine methylations. The parameters may be
CC compared to another set of genetic and/or epigenetic parameters, the
CC differences serving as basis for diagnosis and/or prognosis events which
CC are disadvantageous to patients. The present sequence is one of the
CC 533 genomic sequences derived from tumour suppressor genes and
CC oncogenes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
CX XX
SO Sequence 7667 BP; 2446 A; 63 C; 1493 G; 3665 T; 0 other;
Query Match 4.2%; Score 57.2; DB 22; Length 7667;
Best Local Similarity 60.1%; Pred.No.1.ee=06;
Matches 95; Conservative 0; Mismatches 63; Indels 0; Gaps 0;
OY 76 TAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGC 135
DB 4557 TAGTGGTGTCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTATATTATAATTATATANA 4616
OY 136 AGAACCAATCCTTAACCCCCTTTTPPGACAACACTGSATCATCTTATTTAAGGCCATTT 195
DB 4617 CGAAAGTATTTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4676
OY 196 CTCAGATTGTAAATGAGCACAGAGCATCTACTTCGAAGTTT 233
DB 4677 GTAGTTTATAATGTATATATTATTTGATTAGTAGATTTT 4714
Search completed: February 17, 2003, 09:18:44
Job time : 286.868 secs

```